

Protein - protein search, using sw mode!									
Date:	May 24, 2004, 14:52:18 ; Search time 43 Seconds (without alignments) 45.406 Million cell updates/sec								
on:									
Sequence:	Effect score: 33 1 KSADLKR 7								
oring table:	BLOSUM62 Gapext 0.5								
searched:	1149313 seqs, 278921704 residues								
all number of hits satisfying chosen parameters:	1149313								
minimum DB length:	0								
maximum DB length:	2000000000								
st-processing:	Minimum Match 0% Maximum Match 100% Listing first 45 summaries								
atabase :	Published_Applications_HA: 1: /cgn2_6_ptodata/1/pubpaas/US07_PUBCOMB.pep; 2: /cgn2_6_ptodata/1/pubpaas/BCT_NEW_PUB..pep; 3: /cgn2_6_ptodata/1/pubpaas/US06_N_PUB..pep; 4: /cgn2_6_ptodata/1/pubpaas/US06_PUBCOMB..pep; 5: /cgn2_6_ptodata/1/pubpaas/US07_NEW_PUB..pep; 6: /cgn2_6_ptodata/1/pubpaas/PCTUS_PUBCOMB..pep; 7: /cgn2_6_ptodata/1/pubpaas/US08_NEW_PUB..pep; 8: /cgn2_6_ptodata/1/pubpaas/PUBCOMB..pep; 9: /cgn2_6_ptodata/1/pubpaas/US09_PUBCOMB..pep; 10: /cgn2_6_ptodata/1/pubpaas/US09C_PUBCOMB..pep; 11: /cgn2_6_ptodata/1/pubpaas/US09E_PUBCOMB..pep; 12: /cgn2_6_pcdata/1/pubpaas/US09_N_PUB..pep; 13: /cgn2_6_pcdata/1/pubpaas/US10A_PUBCOMB..pep; 14: /cgn2_6_pcdata/1/pubpaas/US10B_PUBCOMB..pep; 15: /cgn2_6_pcdata/1/pubpaas/US10C_PUBCOMB..pep; 16: /cgn2_6_pcdata/1/pubpaas/US10_NEW_PUB..pep; 17: /cgn2_6_pcdata/1/pubpaas/US60_NEW_PUB..pep; 18: /cgn2_6_pcdata/1/pubpaas/US60_PUBCOMB..pep;								
Prod. No.	is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.								
	SUMMARIES								
	* Query Match Score Length DB ID Description								
Result No.	Score	Match	Length	DB	ID	Description	Query	Match Score	Length;
1	33	100.0	7	13	US-10-080-100-1	Sequence 1, Appli	1	100.0%; Best Local Similarity 100.0%; Matches 7; Conservative 0; Mismatches 0;	DB 13; Index 0; Gaps 0;
2	33	100.0	7	13	US-10-080-100-4	Sequence 2, Appli			
3	33	100.0	35	14	US-10-022-914-22	Sequence 3, Appli			
4	33	100.0	67	10	US-09-500-700-39	Sequence 4, Appli			
5	33	100.0	98	10	US-09-565-700-42	Sequence 5, Appli			
6	33	100.0	105	10	US-09-908-152-B-34	Sequence 6, Appli			
7	33	100.0	189	9	US-10-157-899-A-2	Sequence 7, Appli			
8	33	100.0	414	14	US-10-157-899-A-6	Sequence 8, Appli			
9	33	100.0	414	14	US-10-157-899-A-8	Sequence 9, Appli			
10	33	100.0	414	14	US-10-157-899-A-10	Sequence 10, Appli			
11	33	100.0	438	14	US-10-157-899-A-4	Sequence 11, Appli			
13	33	100.0	438	14	US-10-157-899-A-12	Sequence 12, Appli			
14	33	100.0	438	14	US-10-157-899-A-14	Sequence 13, Appli			
15	33	100.0	438	14	US-10-157-899-A-16	Sequence 14, Appli			
	ALIGNMENTS								
	RESULT 1								
	Sequence 1, Application US/10080100 Publication No. US2002016556A1								
	; GENERAL INFORMATION: ; APPLICANT: Barbas, Carlos ; INVENTOR: Dreier, Birgit ; TITLE OF INVENTION: Zinc Finger Binding Domains for Nucleotide Sequence ANN ; FILE REFERENCE: TSVI 760.0 ; CURRENT FILING DATE: 2002-02-21 ; PRIORITY APPLICATION NUMBER: US/10/080,100 ; PRIORITY FILING DATE: 2001-02-21 ; NUMBER OF SEQ ID NOS: 113 ; SEQ ID NO: 1 ; LENGTH: 7 ; TYPE: PRT ; ORGANISM: Artificial Sequence ; FEATURE: ; OTHER INFORMATION: synthesized								
	US-10-080-100-1								
	RESULT 2								
	Sequence 2, Application US/10080100 Publication No. US2002016556A1								
	; GENERAL INFORMATION: ; APPLICANT: KSAIDLKR 7 ; INVENTOR: KSAIDLKR 7 ; TITLE OF INVENTION: Zinc Finger Binding Domains for Nucleotide Sequence ANN ; FILE REFERENCE: TSVI 760.0 ; CURRENT FILING DATE: 2002-02-21 ; PRIORITY APPLICATION NUMBER: US/09/791,106 ; PRIORITY FILING DATE: 2001-02-21 ; NUMBER OF SEQ ID NOS: 113 ; SEQ ID NO: 1 ; LENGTH: 7 ; TYPE: PRT ; ORGANISM: Artificial Sequence ; FEATURE: ; OTHER INFORMATION: synthesized								
	US-10-080-100-1								

APPLICANT: Barbas, Carlos
 APPLICANT: Dreier, Birgit
 TITLE OF INVENTION: Zinc Finger Binding Domains for Nucleotide Sequence ANN
 FILE REFERENCE: SCRIP1160-4
 CURRENT APPLICATION NUMBER: US/09/500,700
 CURRENT FILING DATE: 2003-01-10
 PRIOR APPLICATION NUMBER: US/08/863,813
 PRIOR FILING DATE: 1997-05-27
 PRIOR APPLICATION NUMBER: US/08/676,318
 PRIOR FILING DATE: 1996-12-30
 PRIOR APPLICATION NUMBER: PCT/US95/00829
 PRIOR FILING DATE: 1995-01-18
 PRIOR APPLICATION NUMBER: US/08/312,604
 PRIOR FILING DATE: 1994-09-28
 PRIOR APPLICATION NUMBER: US/08/183,119
 PRIOR FILING DATE: 1994-01-18
 NUMBER OF SEQ ID NOS: 113
 SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO: 4
 LENGTH: 7
 TYPE: PRT
 ORGANISM: Artificial sequence
 FEATURE:
 OTHER INFORMATION: synthesized
 US-10-080-100-4

Query Match Similarity 100.0%; Score 33; DB 13; Length 7;
 Best Local Similarity 100.0%; Pred. No. 1e-06;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 K5ADLKR 7
 Db 1 K5ADLKR 7

RESULT 3
 US-10-122-934-22
 Sequence 22, Application US/10422934
 Publication No. US20030186841A1
 GENERAL INFORMATION:
 APPLICANT: Barbas, Carlos F., III
 APPLICANT: Rada, Michael
 APPLICANT: Berli, Roger
 TITLE OF INVENTION: LIGAND ACTIVATED TRANSCRIPTIONAL REGULATOR PROTEINS
 FILE REFERENCE: 22908-1227C
 CURRENT APPLICATION NUMBER: US/10/422,934
 CURRENT FILING DATE: 2003-04-23
 PRIOR APPLICATION NUMBER: 09/586,625
 PRIOR FILING DATE: 2000-06-02
 PRIOR APPLICATION NUMBER: 09/433,042
 PRIOR FILING DATE: 1999-10-25
 NUMBER OF SEQ ID NOS: 92
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO: 22
 LENGTH: 35
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Recombinant
 OTHER INFORMATION: molecule
 US-10-122-934-22

Query Match Similarity 100.0%; Score 33; DB 14; Length 35;
 Best Local Similarity 100.0%; Pred. No. 2.7%;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 K5ADLKR 7
 Db 20 K5ADLKR 26

RESULT 4
 US-09-500-700-39
 Sequence 39, Application US/09500700
 Publication No. US20030059767A1
 GENERAL INFORMATION:
 APPLICANT: THE SCRIPPS RESEARCH INSTITUTE
 APPLICANT: BARBAS III, Carlos F.
 APPLICANT: GOTTESFELD, Joel M.
 APPLICANT: WRIGHT, Peter E.

TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES AND METHODS THEREFOR
 FILE REFERENCE: SCRIP1160-4
 CURRENT APPLICATION NUMBER: US/09/500,700
 CURRENT FILING DATE: 2003-01-10
 PRIOR APPLICATION NUMBER: US/08/863,813
 PRIOR FILING DATE: 1997-05-27
 PRIOR APPLICATION NUMBER: US/08/676,318
 PRIOR FILING DATE: 1996-12-30
 PRIOR APPLICATION NUMBER: PCT/US95/00829
 PRIOR FILING DATE: 1995-01-18
 PRIOR APPLICATION NUMBER: US/08/312,604
 PRIOR FILING DATE: 1994-09-28
 PRIOR APPLICATION NUMBER: US/08/183,119
 NUMBER OF SEQ ID NOS: 127
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO: 39
 LENGTH: 67
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Construction of C7 zinc finger
 FEATURE:
 NAME/KEY: MISC_FEATURE
 LOCATION: (67) . (67)
 OTHER INFORMATION: Xaa is Tyr-Ala-Cys-Pro-Val-Glu-Ser-Cys-Arg-Phe-Ser-Lys-Arg-His-Ile-Arg-Leu-Tyr-Gly-Glu-Lys-Pro
 OTHER INFORMATION: Could be repeated 10 times
 US-09-500-700-39

RESULT 5
 US-09-500-700-42
 Sequence 42, Application US/09500700
 Publication No. US20030059767A1
 GENERAL INFORMATION:
 APPLICANT: THE SCRIPPS RESEARCH INSTITUTE
 APPLICANT: BARBAS III, Carlos F.
 APPLICANT: GOTTESFELD, Joel M.
 APPLICANT: WRIGHT, Peter E.

TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES AND METHODS THEREFOR
 FILE REFERENCE: SCRIP1160-4
 CURRENT APPLICATION NUMBER: US/09/500,700
 CURRENT FILING DATE: 2003-01-10
 PRIOR APPLICATION NUMBER: US/08/863,813
 PRIOR FILING DATE: 1997-05-27
 PRIOR APPLICATION NUMBER: US/08/676,318
 PRIOR FILING DATE: 1996-12-30
 PRIOR APPLICATION NUMBER: PCT/US95/00829
 PRIOR FILING DATE: 1995-01-18
 PRIOR APPLICATION NUMBER: US/08/312,604
 PRIOR FILING DATE: 1994-09-28
 PRIOR APPLICATION NUMBER: US/08/183,119
 NUMBER OF SEQ ID NOS: 127
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO: 42
 LENGTH: 98
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-500-700-42

Query Match Similarity 100.0%; Score 33; DB 10; Length 98;
 Best Local Similarity 100.0%; Pred. No. 7.9%;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6
 US-09-765-555-74
 / Sequence 74, Application US/09765555
 / Publication No. US2003003755A1
 / GENERAL INFORMATION:
 / APPLICANT: The Scripps Research Institute
 / TITLE OF INVENTION: Methods and compositions to modulate
 / expression in plants
 / FILE REFERENCE: 27801-20014.40
 / CURRENT APPLICATION NUMBER: US/09/765,555
 / CURRENT FILING DATE: 2002-05-14
 / PRIOR APPLICATION NUMBER: US 09/620,897
 / PRIOR FILING DATE: 2000-01-21
 / PRIOR APPLICATION NUMBER: US 60/177,468
 / PRIOR FILING DATE: 2000-01-21
 / NUMBER OF SEQ ID NOS: 75
 / SOFTWARE: FastSEQ for Windows Version 4.0
 / SEQ ID NO: 74
 / LENGTH: 105
 / TYPE: PRT
 / ORGANISM: Artificial sequence
 / FEATURE:
 / OTHER INFORMATION: Amino acid sequence of 3 finger protein C7
 US-09-765-555-74
 Query Match Score 33; DB 10; Length 105;
 Best Local Similarity 100.0%; Pred. No. 8.5%;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 KSADLKR 7
 Db 22 KSADLKR 28

RESULT 7
 US-09-908-153B-14
 / Sequence 34, Application US/09908153B
 / Patent No. US20020168714A1
 / GENERAL INFORMATION:
 / APPLICANT: Barbas, Carlos F.
 / APPLICANT: Beirer, Roger
 / APPLICANT: Schoofter, Ulrich
 / TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
 / SINGLE-CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
 / TITLE OF INVENTION: SWITCHES
 / FILE REFERENCE: TISRI 725_1
 / CURRENT APPLICATION NUMBER: US/09/908,153B
 / CURRENT FILING DATE: 2001-07-18
 / PRIOR APPLICATION NUMBER: US 09/619,063
 / PRIOR FILING DATE: 2000-07-18
 / NUMBER OF SEQ ID NOS: 53
 / SOFTWARE: FastSEQ for Windows Version 4.0
 / SEQ ID NO: 34
 / LENGTH: 189
 / TYPE: PRT
 / ORGANISM: Unknown
 / FEATURE:
 / OTHER INFORMATION: Synthesized
 / FEATURE:
 / NAME/KEY: VARIANT
 / LOCATION: 1..189
 / OTHER INFORMATION: Xaa = Any Amino Acid

Query Match Score 33; DB 9; Length 189;
 Best Local Similarity 100.0%; Pred. No. 1.6%;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8
 US-10-157-899A-2
 / Sequence 2, Application US/10157899A
 / Publication No. US20030143559A1
 / GENERAL INFORMATION:
 / APPLICANT: Bracken, Kathryn Rene
 / APPLICANT: de los Angeles, Joseph Ernest
 / APPLICANT: Kadan, Michael Joseph
 / APPLICANT: Keander, Gary Michael
 / APPLICANT: Zerry, Dennis
 / TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
 / LIGANDS AND PHARMACEUTICAL COMPOSITIONS
 / CURRENT APPLICATION NUMBER: US/10/157,899A
 / CURRENT FILING DATE: 2002-09-12
 / PRIOR APPLICATION NUMBER: US 60/294,839
 / PRIOR FILING DATE: 2001-05-31
 / NUMBER OF SEQ ID NOS: 55
 / SOFTWARE: PatentIn version 3.1
 / SEQ ID NO: 2
 / LENGTH: 414
 / TYPE: PRT
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: Fusion of the human-Estrogen-receptor-ligand-binding-domain and
 / zinc finger array(C7)
 US-10-157-899A-2
 Query Match Score 33; DB 14; Length 414;
 Best Local Similarity 100.0%; Pred. No. 36%;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 KSADLKR 7
 Db 22 KSADLKR 28

RESULT 9
 US-10-157-899A-6
 / Sequence 6, Application US/10157899A
 / Publication No. US20030143559A1
 / GENERAL INFORMATION:
 / APPLICANT: Bracken, Kathryn Rene
 / APPLICANT: de los Angeles, Joseph Ernest
 / APPLICANT: Huang, Ying
 / APPLICANT: Kadan, Michael Joseph
 / APPLICANT: Keander, Gary Michael
 / APPLICANT: Zerry, Dennis
 / TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
 / LIGANDS AND PHARMACEUTICAL COMPOSITIONS
 / CURRENT APPLICATION NUMBER: US/10/157,899A
 / CURRENT FILING DATE: 2002-09-12
 / PRIOR APPLICATION NUMBER: US 60/294,839
 / PRIOR FILING DATE: 2001-05-31
 / NUMBER OF SEQ ID NOS: 55
 / SOFTWARE: PatentIn version 3.1
 / SEQ ID NO: 6
 / LENGTH: 414
 / TYPE: PRT
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
 / and a zinc finger array(C7)
 US-10-157-899A-6

Query Match 100.0%; Score 33; DB 14; Length 414;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 K\$ADLKR 7
 Db 22 K\$ADLKR 28

RESULT 10
 US-10-157-899A-8
 Sequence 8, Application US/10157899A
 Publication No. US20030143559A1
 GENERAL INFORMATION:
 APPLICANT: Bracken, Kathryn Rene
 APPLICANT: Huang, Ying
 APPLICANT: Kadan, Michael Joseph
 APPLICANT: Zarby, Dennis
 APPLICANT: Ksander, Gary Michael
 TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
 TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
 FILE REFERENCE: 4-32018A
 CURRENT APPLICATION NUMBER: US/10/157,899A
 CURRENT FILING DATE: 2002-09-12
 PRIOR APPLICATION NUMBER: US 60/294,839
 PRIOR FILING DATE: 2001-05-31
 NUMBER OF SEQ ID NOS: 55
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO: 8
 LENGTH: 414
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
 US-10-157-899A-8
 Query Match 100.0%; Score 33; DB 14; Length 414;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 K\$ADLKR 7
 Db 22 K\$ADLKR 28

RESULT 11
 US-10-157-899A-10
 Sequence 10, Application US/10157899A
 Publication No. US20030143559A1
 GENERAL INFORMATION:
 APPLICANT: Bracken, Kathryn Rene
 APPLICANT: Huang, Ying
 APPLICANT: Kadan, Michael Joseph
 APPLICANT: Zarby, Dennis
 APPLICANT: Ksander, Gary Michael
 TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
 TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
 FILE REFERENCE: 4-32018A
 CURRENT APPLICATION NUMBER: US/10/157,899A
 CURRENT FILING DATE: 2002-09-12
 PRIOR APPLICATION NUMBER: US 60/294,839
 PRIOR FILING DATE: 2001-05-31
 NUMBER OF SEQ ID NOS: 55
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO: 10
 LENGTH: 414
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain

OTHER INFORMATION: and a zinc finger array (C7)
 US-10-157-899A-10
 Query Match 100.0%; Score 33; DB 14; Length 414;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 K\$ADLKR 7
 Db 22 K\$ADLKR 28

RESULT 12
 US-10-157-899A-4
 Sequence 4, Application US/10157899A
 Publication No. US20030143559A1
 GENERAL INFORMATION:
 APPLICANT: Bracken, Kathryn Rene
 APPLICANT: Huang, Ying
 APPLICANT: Kadan, Michael Joseph
 APPLICANT: Zarby, Dennis
 APPLICANT: Ksander, Gary Michael
 TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
 TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
 FILE REFERENCE: 4-32018A
 CURRENT APPLICATION NUMBER: US/10/157,899A
 CURRENT FILING DATE: 2002-09-12
 PRIOR APPLICATION NUMBER: US 60/294,839
 PRIOR FILING DATE: 2001-05-31
 NUMBER OF SEQ ID NOS: 55
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO: 4
 LENGTH: 438
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Fusion of the human-Estrogen-receptor-ligand-binding-domain and
 OTHER INFORMATION: Zinc finger array (C7)
 US-10-157-899A-4
 Query Match 100.0%; Score 33; DB 14; Length 438;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 K\$ADLKR 7
 Db 22 K\$ADLKR 28

RESULT 13
 US-10-157-899A-12
 Sequence 12, Application US/10157899A
 Publication No. US20030143559A1
 GENERAL INFORMATION:
 APPLICANT: Bracken, Kathryn Rene
 APPLICANT: Huang, Ying
 APPLICANT: Kadan, Michael Joseph
 APPLICANT: Zarby, Dennis
 APPLICANT: Ksander, Gary Michael
 TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
 TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
 FILE REFERENCE: 4-32018A
 CURRENT APPLICATION NUMBER: US/10/157,899A
 CURRENT FILING DATE: 2002-09-12
 PRIOR APPLICATION NUMBER: US 60/294,839
 PRIOR FILING DATE: 2001-05-31
 NUMBER OF SEQ ID NOS: 55
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO: 12
 LENGTH: 438
 TYPE: PRT

```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
US-10-157-899A-12

Query Match      100.0%; Score 33; DB 14; Length 438;
Best Local Similarity 100.0%; Prd. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      22 KSADLKR 28

RESULT 14
US-10-157-899A-14
Sequence 14, Application US/10157899A
Publication No. US2003014359A1
GENERAL INFORMATION:
APPLICANT: Bracken, Kathryn Rene
APPLICANT: de Los Angeles, Joseph Ernest
APPLICANT: Huang, Ying
APPLICANT: Kadan, Michael Joseph
APPLICANT: Krander, Gary Michael
APPLICANT: Zerby, Dennis
TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
FILE REFERENCE: 4-320:8A
CURRENT APPLICATION NUMBER: US/10/157,899A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 60/294,839
PRIOR FILING DATE: 2001-05-31
NUMBER OF SEQ ID NOS: 55
SOFTWARE: Patentin version 3.1
SEQ ID NO 14
LENGTH: 438
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
OTHER INFORMATION: and a zinc finger array(C7)
US-10-157-899A-14

Query Match      100.0%; Score 33; DB 14; Length 438;
Best Local Similarity 100.0%; Prd. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      22 KSADLKR 28

```

```

RESULT 15
US-10-157-899A-16
Sequence 16, Application US/10157899A
Publication No. US2003014359A1
GENERAL INFORMATION:
APPLICANT: Bracken, Kathryn Rene
APPLICANT: de Los Angeles, Joseph Ernest
APPLICANT: Huang, Ying
APPLICANT: Kadan, Michael Joseph
APPLICANT: Krander, Gary Michael
APPLICANT: Zerby, Dennis
TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
FILE REFERENCE: 4-320:8A
CURRENT APPLICATION NUMBER: US/10/157,899A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 60/294,839
PRIOR FILING DATE: 2001-05-31
NUMBER OF SEQ ID NOS: 55
SOFTWARE: Patentin version 3.1

```

```

; SEQ ID NO 16
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
US-10-157-899A-16

Query Match      100.0%; Score 33; DB 14; Length 438;
Best Local Similarity 100.0%; Prd. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      22 KSADLKR 28

Search completed: May 24, 2004, 14:58:02
Job time : 43 secs

```

achieved through specific contacts from side chains of amino acid residues in the alpha helix. Each zinc finger can recognise a subsite of 3 bp in target DNA. Covalent linkage of multiple zinc finger domains allows the recognition of extended contiguous asymmetric DNA sequences. For example, a synthetic polydactyl protein containing six zinc finger domains can recognise an 18 bp sequence, and such proteins are potentially highly gene-specific. The novel nucleotide-binding zinc finger proteins may therefore be used in the development of artificial gene-specific transcriptional regulators. Such transcriptional switches may be used to regulate the expression of oncogenes such as erbB-2, overexpression of which is involved in malignant transformation. The proteins are therefore useful in the treatment of cancers, and may also be used to activate genes involved in fighting diseases, and to treat viral infections by inhibiting the synthesis of viral gene products. They may be used in DNA-based diagnostic applications. The proteins may also be used in producing functional gene products. The proteins may also be heterozygous transgenic animals. Proteins of the invention can discriminate between sequences which have a single base difference. This is manifested in a 2100-fold decrease in affinity for the variant sequence. Gene activation and repression can be achieved by targeting within the gene transcript, suggesting that information obtained from expressed sequence tags may be sufficient for the construction of gene switches. Sequences AB02876-302869 represent zinc finger alpha helix phage library peptides disclosed in the invention.

SQ Sequence 7 AA;

Query	Match 100.0%	Score 33;	DB 3;	Length 7;
	Best Local Similarity 100.0%;	Pred. No. 1.4e+06;		
	Matches 7;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1 K\$ADLKR 7			
Db	1 K\$ADLKR 7			

RESULT 2

ABP53161

ID ABP53161 standard; peptide; 7 AA.

XX

AC ABP53161;

XX

DT 12-NOV-2002 (first entry)

XX

DE Mouse transcription factor Zif268 zinc finger helix peptide SEQ ID NO:4

XX

PF Zinc finger binding domain; zinc finger nucleotide binding peptide;

XX

PT zinc finger; Gene expression; modulation; promoter; viral; Lentivirus;

XX

PR human T-cell lymphocytic virus; HTLV; human immunodeficiency virus; HIV.

XX

OS Mus sp.

XX

PN WO200266640-A2.

XX

PD 29-AUG-2002.

XX

DR WPI; 2002-674941/72.

XX

PT New polypeptides comprising 2 to 10 zinc finger-nucleotide binding

PT peptides, useful for modulating gene expression in promoters from viral

PT groups, e.g. Lentivirus group, such as human T-cell lymphocytic virus; HTLV-1 and 2, or HIV-1 and 2.

XX

PS Example 1; Fig 1; 48pp; English.

XX

CC The present invention describes a polypeptide comprising 2 to 12 zinc

CC finger-nucleotide binding peptides in which at least one contains a

CC nucleotide binding region. Also described: (1) an isolated and purified

CC polynucleotide encoding the polypeptide cited above; (2) a process of

CC vector containing the polynucleotide in (1); and (3) a process of

CC regulating expression of a nucleotide sequence that contains the sequence

CC S1 comprising exposing the nucleotide sequence to the polypeptide cited

CC above. (S1) is (5'-ANN)n-3', where n = any integer from 2-12; and N = A,

CC C, G, or T. The polypeptide is useful for modulating gene expression in

CC

XX

CC CC CC CC

Sequence Comparison Report							
Query Sequence		Database Sequence					
Qb	1 KSAIDLKR 7	1 KSAIDLKR 7	Score 33;	DB 4;	Length 35;	0;	0;
AAE02027	ARE02027 Standard; peptide; 35 AA.	XX	Best Local Similarity	100.0%	Score 33;	Pred. No. 5;	0;
AAE02027:	31-JUL-2001 (first entry)	XX	Best Local Similarity	100.0%	Score 33;	DB 2;	Length 64;
RESULT 6							
AAE02027	Murine Zif C7 ZFP modified fragment #1.	XX	Query Match	100.0%	Score 33;	DB 4;	Length 35;
RESULT 7							
AAE02027	Fusion protein; nucleotide-binding domain; NBD; ligand-binding domain; LBD; transcription regulating domain; TRD; zinc finger protein; ZFP; ligand-activated transcriptional regulator; gene regulation; Gene therapy; cell proliferative disorder; cancer; psoriasis; pemphigus vulgaris; Behcet's syndrome; lipid histiocytosis; Zif-C7; murine.	XX	Best Local Similarity	100.0%	Score 33;	Pred. No. 5;	0;
RESULT 8							
AAE02027	Mus sp. Synthetic.	XX	Query Match	100.0%	Score 33;	DB 4;	Length 35;
RESULT 9							
AAE02027	New fusion protein containing nucleotide-binding and ligand-binding domains, useful e.g. in gene therapy of cancer, provides ligand-activated control of gene expression.	XX	Best Local Similarity	100.0%	Score 33;	Pred. No. 5;	0;
RESULT 10							
AAE02027	Disclosure: Page 49; 218sp; English.	XX	Query Match	100.0%	Score 33;	DB 4;	Length 35;
RESULT 11							
AAE02027	The invention relates to fusion protein comprising a nucleotide-binding domain (NBD), a ligand-binding domain (LBD) of an intracellular receptor (ICR) and a transcription regulating domain (TRD). NBD is a polydactyl zinc finger protein (ZFP), or a modular part of it, that interacts specifically with a contiguous sequence of at least 2 nucleotides. The fusion protein functions as a ligand-activated transcriptional regulator. The fusion protein and the nucleic acid encoding it, are used to regulate gene expression, particularly in gene therapy for treating malignant cell proliferative diseases (e.g. colon cancer, prostate cancer, renal-cell carcinoma) and non-malignant cell proliferative diseases (e.g. psoriasis, pemphigus vulgaris, Behcet's syndrome and lipid histiocytosis). The fusion protein and its DNA are also useful for treating diseases caused by viruses in humans/plants, generic and/or acquired diseases. The fusion protein can be designed to target any selected gene (endogenous or exogenous), and can be made to have different selectivity or specificity for endogenous or exogenous ligands. The present sequence is murine Zif-C7 ZFP fragment comprising modified finger 1 which has altered DNA binding specificity	XX	Best Local Similarity	100.0%	Score 33;	Pred. No. 5;	0;
RESULT 12							
AAE02027	Sequence 35 AA;	XX	Query Match	100.0%	Score 33;	DB 4;	Length 35;
RESULT 13							
AAE02027	Barbas CF, Gottsfield JM, Wright PE; WPI; 1995-263862/34.	XX	Best Local Similarity	100.0%	Score 33;	Pred. No. 5;	0;
RESULT 14							
AAE02027	New zinc finger-nucleotide binding polypeptide(s) - used for modulating the function of cellular nucleotide sequences, partic. for treating cell proliferative disorders.	XX	Query Match	100.0%	Score 33;	DB 4;	Length 64 AA;
RESULT 15							
AAE02027	Example 13: Page 86; 135pp; English.	XX	Best Local Similarity	100.0%	Score 33;	Pred. No. 5;	0;
RESULT 16							
AAE02027	Mutagenesis of the Zif268 protein produced a variant (designated C7) with an enhanced affinity for the GCG-GCC-GCG target sequence. Proteins containing 2 to 12 copies of the C7 finger, linked via the consensus linker TGExP, were constructed and were shown to have specificity for their predicted targets. The present sequence is that of the protein containing 2 copies of C7. Fingers utilised in the multifinger proteins need not be identical and may be mixed to produce proteins which recognise a desired target sequence. They may also be used with leucine zippers to produce proteins with extended sequence recognition. The proteins are useful for suppressing transcription and therefore for treating proliferative disorders	XX	Query Match	100.0%	Score 33;	DB 4;	Length 64;

RESULT 8							
Best Local Similarity		Pred. No. 9 6;		Mismatches 0;		Indels 0; Gaps 0;	
Matches	7;	Conservative					
Qy	1 KSAIDLKR 7						
Dbx	20 KSAIDLKR 26						
 RESULT 9							
ID	AAW84390	standard; protein; 64 AA.					
XX	AAW84390;						
AC	AAW84390;						
CC	AAW84390;						
DT	AAW84390;						
XX	AAW84390;						
DE	Zig268 C7 finger sequence.						
XX	Zig268 three finger protein; zinc finger; nucleotide-binding protein; cell proliferative disorder; gene therapy; cancer; psoriasis; Bemphigus vulgaris; Bechet's syndrome; lipid histiocytosis; human immune deficiency virus; HIV; viral infection; transgenic plant.						
DS	Synthetic.						
XX							
PN	WO9854311-A1.						
XX							
PD	03-DEC-1998.						
PP	27-MAY-1998;	98WO-US010801.					
XX							
PR	27-MAY-1997;	97US-00863813.					
XX							
PA	(SCRIPPS RES INST.						
XX							
PPI	Barbas CF, Gottesfeld JM, Wright PE;						
XX							
DR	WPI; 1999-059831/05.						
XX							
PPT	New zinc finger nucleotide-binding protein variant that modulates selected nucleotide sequence - used for treatment of proliferative and viral diseases by gene therapy, and can be made selective for any target sequence.						
PP	Example 13; Page 89; 158pp; English.						
PP	Sequence 64 AA;						
XX	Query Match Similarity 100.0%; Score 33; DB 2; Length 64;						
CC	Best Local Similarity 100.0%; Pred. No. 9 6;	Mismatches 0;	Indels 0;	Gaps 0;			
CC	Matches 7;	Conservative					
CC	1 KSAIDLKR 7						
CC	20 KSAIDLKR 26						

gene therapy; cancer; psoriasis; *pemphigus vulgaris*; Bechet's syndrome;
 lipid histiocytosis; human immune deficiency virus; HIV; viral infection;
 transgenic plant; C7.
 XX
 KW OS
 KW Synthetic.
 KW
 Key Location/Qualifiers
 PH Misc-difference 91 /note= "encoded by CAC"
 FT Misc-difference 98 /note= "encoded by ACT"
 FT PN WO954311-A1.
 XX PD 03 -DBC-1998.
 XX PP 27-MAY-1998; 98WO-US510801.
 XX PR 27-MAY-1997; 97US-00363813.
 XX PA (Scripps Res Inst.
 XX PI Barbas CF, Gottesfeld JM, Wright PE;
 XXX WPI; 1999-053831/05.
 DR N-PDB; AAX03253.
 XX
 New zinc finger nucleotide-binding protein variant that modulates selected nucleotide sequence - used for treatment of proliferative and viral diseases by gene therapy, and can be made selective for any target sequence.
 Example 13; Fig 15; 158pp; English.
 The present sequence represents a three zinc finger C7 protein, made in the course of the invention. The specification describes zinc finger nucleotide-binding protein variants with at least two zinc finger modules that bind to a cellular nucleotide sequence and modulate its function. Zinc finger proteins, and compositions containing them, are used to increase or reduce transcription of a gene linked to the cellular nucleotide sequence. The proteins are used specifically for treating or preventing cell proliferative disorders (in humans, animals or plants, including those induced by viruses). Particularly where expressed from nucleic acid by gene therapy (including ex vivo methods). Typical diseases that can be treated are many forms of cancer, psoriasis, *pemphigus vulgaris*, Bechet's syndrome and lipid histiocytosis, also treatment of human immune deficiency virus (HIV) and other viral infections, and production of transgenic plants resistant to bacterial and viral diseases.

```

Sequence 30 AA;
Query Match          100.0%; Score 33; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 15;
Matches    7; Conservative 0; Mismatches 0; Indels 0;
QY          1 K$ADLKR 7

```

RESULT 12
AAW84455
TTT

RESULT 11
 AAAE06026
 AAE06026 standard; protein; 105 AA.
 XX
 AC
 XX
 DT
 24-SEP-2001 (first entry)
 Three finger protein (C7).
 XX
 Zinc finger protein; ZFP; gene expression; metabolic pathway regulator;
 modular; plant technology; architecture; three finger protein C7

PN WO9854311-A1.
 XX
 PD 03-DEC-1998.
 XX
 XX 27-MAY-1998; 98W0-US010801.
 XX
 PR 27-MAY-1997; 97US-00863813.
 XX
 PA (SCRI) SCRIPPS RES INST.
 XX
 PI Barbas CF, Gottesfeld JM, Wright PE;
 XX
 DR 1999-059831/05.
 XX
 PT New zinc finger nucleotide-binding protein variant that modulates selected nucleotide sequence - used for treatment of proliferative and viral diseases by gene therapy, and can be made selective for any target sequence.
 XX
 PS Disclosure; Page 53; 158pp; English.
 XX
 CC The present sequence can be used to construct C7 zinc fingers, in the course of the invention. The specification describes zinc finger nucleotide-binding protein variants with at least two zinc finger modules that bind to a cellular nucleotide sequence and modulate its function.
 CC Zinc finger proteins, and compositions containing them, are used to increase or reduce transcription of a gene linked to the cellular nucleotide sequence. The proteins are used specifically for treating or preventing cell proliferative disorders (in humans, animals or plants, including those induced by viruses), particularly where expressed from nucleic acid by gene therapy (including ex vivo methods). Typical diseases that can be treated are many forms of cancer, psoriasis, Emphysema vulgaris, Bechet's syndrome and lipid histiocytosis, also treatment of human immune deficiency virus (HIV) and other viral infections, and production of transgenic plants resistant to bacterial and viral diseases
 XX Sequence 109 AA;
 SQ Query Match 100.0%; Score 33; DB 2; Length 109;
 Best Local Similarity 100.0%; Prod. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 K\$ADLKR 7
 Db 20 K\$ADLKR 26
 XX
 RESULT 13
 AAB47900 AAB47900 standard; protein; 189 AA.
 XX
 ID AAB47900
 AC AAB47900;
 XX DT 16-MAY-2002 (first entry)
 XX DE Zinc finger DNA binding domain, 2C7.
 XX KW Polymerase chain reaction; primer; amplify; PCR; reporter;
 KW nucleotide gene switch; ligand binding domain; LBD; transcription;
 KW nuclear hormone receptor; gene therapy; regulation; expression.
 XX OS Synthetic.
 XX
 Key Location/Qualifiers
 PH Misc-difference 1 /note= "Unspecified amino acid encoded by NNG"
 FT FT Misc-difference 189 /note= "Unspecified amino acid encoded by CNN"
 FT FT
 XX PN WO200206461-A2.
 XX
 PD 24-JAN-2002.

XX 16-JUL-2001; 2001WO-EPO08190.
 PP XX
 PR XX
 XX 18-JUL-2000; 2000US-00619063.
 PA (NOVS) NOVARTIS AG.
 PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES NEH.
 PA (SCRI) SCRIPPS RES INST.
 XX
 PI Beerli R, Schopfer U, Barbas CF;
 XX
 DR 2002-179793/23.
 DR N-PSDB; AAI24745.

XX Single chain, monomeric polypeptide gene switches for regulating gene function, comprise two ligand binding domains derived from nuclear hormone receptors operatively linked to one functional domain.
 XX
 PS Disclosure; Fig 5; 61pp; English.
 XX
 CC This sequence represents the zinc finger DNA binding domain 2C7. This protein could be used in the construction of a non-naturally occurring polypeptide gene switch comprising two ligand binding domains (LBDS) derived from nuclear hormone receptors operatively linked to a first functional domain. The gene switch is useful for regulating the function of a target nucleotide that contains a defined sequence, by exposing the target nucleotide to the polypeptide gene switch in the presence of a ligand that binds one of the LBDS of the polypeptide, where the DNA binding domain of the polypeptide binds the defined sequence, or the functional domain of the polypeptide alters the function of the target nucleotide. The gene switch is also useful in the field of gene therapy and as a regulator of gene expression or transcription. The advantage of the gene switches of the invention over existing gene switches is the need for only a single molecular switch and a single expression vector for production of that switch.
 XX
 SQ Sequence 189 AA;
 Query Match 100.0%; Score 33; DB 5; Length 189;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 K\$ADLKR 7
 Db 22 K\$ADLKR 28
 XX
 RESULT 14
 ABU66293 ABU66293 standard; protein; 189 AA.
 ID XX
 AC ABU66293;
 XX DT 23-MAY-2003 (first entry)
 DE Zinc finger binding domain 2C7.
 KW Zinc finger binding domain; gene switch; gene regulation.
 XX
 Unidentified.
 OS
 XX PD 14-NOV-2002.
 XX
 PR 18-JUL-2001; 2001US-00908153.
 XX
 PR 18-JUL-2000; 2000US-00325747.
 XX
 PA (SCRI) SCRIPPS RES INST.
 XX
 PI Barbas CF, Beerli R, Schopfer U;

DR WPI; 2003-328405/31.
 DR N-ISDB; ACA02219.
 XX Novel polypeptide gene switch useful for regulating gene function,
 PT comprises two ligand binding domains derived from nuclear hormone
 PR receptors operatively linked to a functional domain.
 XX Disclosure; Fig 5; 31pp; English.

CC The invention relates to a non-naturally occurring polypeptide (or
 CC polypeptide gene switch) comprising two ligand binding domains derived
 CC from nuclear hormone receptors operatively linked to a first functional
 CC domain. The polypeptide is useful for regulating the function of a target
 CC nucleotide that contains a defined sequence, by exposing the target
 CC nucleotide to the polypeptide in the presence of a ligand that binds one
 CC of the ligand binding domains of the polypeptide, where the DNA binding
 CC domain of the polypeptide binds the defined sequence or alters the produced
 CC function of the target nucleotide. The gene switches can be produced
 CC using a single molecular switch and a single expression vector. The
 CC present sequence represents the amino acid sequence of a zinc finger
 CC binding domain.

XX Sequence 189 AA;
 Query Match 100.0%; Score 33; DB 6; Length 189;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 K5ADLKR 7
 Db 22 K5ADLKR 28

RESULT 15
 AAR88802 ID AAR88802 standard; protein; 354 AA.
 XX AC
 XX DT 19-APR-1996 (First entry)
 DE 11 copies of zinc finger variant C7 fused via consensus linkers.
 XX zinc finger; variant; fusion protein; consensus linker;
 KW multifinger protein; cancer; therapy; virus infection;
 KW transcription-activator; transcription-inhibitor; plant disease.
 XX OS Synthetic.

XX FH Key
 FT Domain 7..30
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"
 FT Region 31..35
 FT /label= "consensus_linker"
 FT 36..59
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"
 FT Region 60..64
 FT /label= "consensus_linker"
 FT 65..88
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"
 FT Region 89..93
 FT /label= "consensus_linker"
 FT 94..117
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"
 FT Region 118..122
 FT /label= "consensus_linker"
 FT 123..146
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"

FT Region 147..151
 FT /label= "consensus_linker"
 FT 152..175
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"
 FT 176..180
 FT /label= "consensus_linker"
 FT 181..204
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"
 FT 205..209
 FT /label= "consensus_linker"
 FT 210..233
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"
 FT 234..238
 FT /label= "consensus_linker"
 FT 235..262
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"
 FT 263..267
 FT /label= "consensus_linker"
 FT 268..291
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"
 FT 292..296
 FT /label= "consensus_linker"
 FT 297..320
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"
 FT 321..325
 FT /label= "consensus_linker"
 FT 326..349
 FT /label= "zinc_finger_C7"
 FT /note= "binds_GCG-GCG-GCG target sequence"
 FT 350..354
 FT /label= "consensus_linker"
 FT /note= "sequence_of terminal linker is subject to change
 as it is not involved in linking two fingers together."
 XX WO9519431-A1.
 XX PD 20-JUL-1995.
 XX PP 18-JAN-1995;
 XX PR 18-JAN-1994;
 XX PR 28-SEP-1994;
 XX PA (SCRI) SCRIPPS RES INST.
 XX Barbas CF, Gottesfeld JM, Wright PE;
 XX PR 1995-263882/34.
 XX New zinc finger-nucleotide binding polypeptide(s) - used for modulating
 PT the function of cellular nucleotide sequences, partic. for treating cell
 PR proliferative disorders.
 XX Example 13; Page 86; 135pp; English.
 XX Mutagenesis of the Zif268 protein produced a variant (designated C7) with
 CC an enhanced affinity for the GCG-GCG-GCG target sequence. Proteins
 CC containing 2 to 12 copies of the C7 finger, linked via the consensus
 CC linker TGEKP, were constructed and were shown to have specificity for
 CC their predicted targets. The present sequence is that of the protein
 CC containing 12 copies of C7. Fingers utilised in the multifinger proteins
 CC need not be identical and may be mixed to produce proteins which
 CC recognise a desired target sequence. They may also be used with leucine
 CC zippers to produce proteins with extended sequence recognition. The
 CC proteins are useful for suppressing transcription and therefore for
 CC treating proliferative disorders

```
SQ Sequence 354 AA;
Query Match 100.0%; Score 33; DS 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 K$ADLKR 7
Db 20 K$ADLKR 26
```

Search completed: May 24, 2004, 14:50:58
Job time : 56 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 14:49:53 ; Search time 22 Seconds

(without alignments)
 16.456 Million cell updates/sec

Title: US-09-807-665-41

Perfect score: 33

Sequence: 1 KSADLKR 7

Scoring table: BLOSUM62

Gappen 0.0 , Gapext 0.5

Searched: 389414 seqs, 51625571 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/2/iaa/5A_COMB.PEP:*

2: /cgn2_6/prodata/2/iaa/5B_COMB.PEP:*

3: /cgn2_6/prodata/2/iaa/6A_COMB.PEP:*

4: /cgn2_6/prodata/2/iaa/6B_COMB.PEP:*

5: /cgn2_6/prodata/2/iaa/BCTUS_COMB.PEP:*

6: /cgn2_6/prodata/2/iaa/backfile1.pep:*

ALIGNMENTS

RESULT 1
 US-09-173-941-41
 / Sequence 41, Application US/09173941
 / Patent No. 6140081

/ GENERAL INFORMATION:

/ TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN

/ FILE REFERENCE: NOV0815

/ CURRENT APPLICATION NUMBER: US/09/173, 941

/ CURRENT FILING DATE: 1998-10-16

/ NUMBER OF SEQ ID NOS: 120

/ SOFTWARE: PatentIn Ver. 2.1

/ SEQ ID NO: 41

/ LENGTH: 7

/ TYPE: PRT
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: Description of Artificial Sequence: nucleotide
 / OTHER INFORMATION: Codon binding sequence

US-09-173-941-41

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	7	US-09-173-941-41	Sequence 41, Appl
2	33	100.0	7	US-09-173-941-41	Sequence 41, Appl
3	33	100.0	64	US-09-173-941-41	Sequence 39, Appl
4	33	100.0	76	US-09-173-941-41	Sequence 39, Appl
5	33	100.0	98	US-09-173-941-42	Sequence 42, Appl
6	33	100.0	98	US-09-173-941-42	Sequence 42, Appl
7	29	87.9	7	US-09-173-941-44	Sequence 44, Appl
8	29	87.9	7	US-09-173-941-44	Sequence 44, Appl
9	29	87.9	149	US-09-173-941-44	Sequence 44, Appl
10	29	87.9	251	US-09-173-941-44	Sequence 44, Appl
11	28	84.8	73	US-09-173-941-44	Sequence 44, Appl
12	28	84.8	156	US-09-173-941-44	Sequence 44, Appl
13	28	84.8	179	US-09-173-941-44	Sequence 44, Appl
14	28	84.8	206	US-09-173-941-44	Sequence 44, Appl
15	28	84.8	423	US-09-173-941-44	Sequence 44, Appl
16	28	84.8	652	US-09-173-941-44	Sequence 44, Appl
17	28	84.8	1297	US-09-173-941-44	Sequence 44, Appl
18	28	84.8	1704	US-09-173-941-44	Sequence 44, Appl
19	28	84.8	1711	US-09-173-941-44	Sequence 44, Appl
20	28	84.8	1711	US-09-173-941-44	Sequence 44, Appl
21	28	84.8	1711	US-09-173-941-44	Sequence 44, Appl
22	27	81.3	180	US-09-173-941-44	Sequence 44, Appl
23	27	81.3	198	US-09-173-941-44	Sequence 44, Appl
24	27	81.3	661	US-09-173-941-44	Sequence 44, Appl
25	26	78.3	7	US-09-173-941-44	Sequence 44, Appl
26	26	78.8	68	US-09-173-941-44	Sequence 44, Appl
27	26	78.8	68	US-09-173-941-44	Sequence 44, Appl

/ LENGTH: 7

/ TYPE: PRT

/ ORGANISM: Artificial Sequence

FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:phage display
; OTHER INFORMATION: selected and mutagenized

US-09-494-190-41

Query Match Score 33; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7
Db 1 KSADLKR 7

RESULT 3
US-08-676-318A-39
Sequence 39, Application US/08676318A

GENERAL INFORMATION:
Patent No. 6242568

APPLICANT: THE SCRIPPS RESEARCH INSTITUTE
TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES AND METHODS THEREFOR
NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,318A

FILING DATE: 30-DEC-1996
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: Haille, Lisa A., Ph.D.
REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 08401/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099

INFORMATION FOR SEQ ID NO: 39:

SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

FEATURE:
NAME/KEY: Peptide
LOCATION: 1..64

US-08-676-318A-39

Query Match Score 33; DB 3; Length 64;
Best Local Similarity 100.0%; Pred. No. 1e+09;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7
Db 20 KSADLKR 26

RESULT 4
US-08-863-813A-39
Sequence 39, Application US/08863813A

Patent No. 6140466

GENERAL INFORMATION:
APPLICANT: Barbas III, Carlos F.
APPLICANT: Wright, Peter E.
TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES
NUMBER OF SEQUENCES: 62

CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA

APPLICANT: Gottsfield, Joel M.
APPLICANT: Wright, Peter E.
TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES
TITLE OF INVENTION: AND METHODS THEREFOR
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/863,813A
FILING DATE: 27-MAY-1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/676,318
FILING DATE: 18-JUL-1996
APPLICATION NUMBER: 08/183,119
FILING DATE: 18-JAN-1996
APPLICATION NUMBER: US95/00829
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Haille, Lisa A., Ph.D.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 08401/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5099
TELEFAX: 619-678-5099

INFORMATION FOR SEQ ID NO: 39:

SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

FEATURE:
LOCATION: 76..76

OTHER INFORMATION: where Xaa at position 76 is Tyr-Ala-Cys-Pro-

OTHER INFORMATION: Val-Glu-Ser-Cys-Hsp-Arg-Arg-Phe-Ser-Lys-Ser-Ala-Asp-Leu-Lys-
OTHER INFORMATION: Thr-Gly-Glu-Lys-Pro-Net-Lys-Ile-Leu-Glu-Pro-repeated 10 time

OTHER INFORMATION: some Xaa's may be missing

US-08-863-813A-39

Query Match Score 33; DB 3; Length 76;
Best Local Similarity 100.0%; Pred. No. 2,2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7
Db 20 KSADLKR 26

RESULT 5
US-08-863-813A-42
Sequence 42, Application US/08863813A
Patent No. 6140466

GENERAL INFORMATION:
APPLICANT: Barbas III, Carlos F.
APPLICANT: Wright, Peter E.
TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES
NUMBER OF SEQUENCES: 62

CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA

COUNTRY: USA
 ZIP: 92037
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: FastSEQ for Windows Version 2.0b
 CURRENT APPLICATION NUMBER: US/08/863,813A
 FILING DATE: 27-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/676,318
 FILING DATE: 18-JUL-1996
 APPLICATION NUMBER: US/08/183,119
 FILING DATE: 18-JUN-1996
 APPLICATION NUMBER: US/95/00829
 FILING DATE: 18-JAN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Haile, Lisa A., Ph.D.
 TELEPHONE: 619/678-5070
 REGISTRATION NUMBER: 38,347
 REFERENCE DOCKET NUMBER: 08401/0100001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619/678-0099
 REGISTRATION NUMBER: 42:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 98 amino acids
 TOPOLogy: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 US-08-863-813A-42

Query Match 100.0%; Score 33; DB 3; Length 98;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 7; Conservative 0; Mismatches 0;
 Indels 0; Gaps 0;

Qy 1 K\$ADLKR 7
 Db 20 K\$ADLKR 26

RESULT 6
 US-08-676-318A-42
 Sequence 42, Application US/08676318A
 GENERAL INFORMATION:
 APPLICANT: THE SCRIPPS RESEARCH INSTITUTE
 TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES AND METHODS THEREFOR
 NUMBER OF SEQUENCES: 44
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson, P.C.
 STREET: 4225 Executive Square, Suite 1400
 CITY: La Jolla
 STATE: California
 ZIP: 92037
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/676,318A
 FILING DATE: 30-DEC-1996
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Haile, Lisa A., Ph.D.
 REGISTRATION NUMBER: 38,347
 REFERENCE DOCKET NUMBER: 08401/0020001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-678-5070

TELEFAX: 619-678-5099
 INFORMATION FOR SEQ ID NO: 42:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 98 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-676-318A-42

Query Match 100.0%; Score 33; DB 3; Length 98;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 7; Conservative 0; Mismatches 0;
 Indels 0; Gaps 0;

Qy 1 K\$ADLKR 7
 Db 20 K\$ADLKR 26

RESULT 7
 US-09-173-941-44
 Sequence 44, Application US/09173941
 GENERAL INFORMATION:
 APPLICANT: BARBAS, Carlos F.
 PATENT NO. 6140081
 TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
 FILE REFERENCE: NOVO081S
 CURRENT APPLICATION NUMBER: US/09/173,941
 CURRENT FILING DATE: 1998-10-16
 NUMBER OF SEQ ID NOS: 120
 SEQ ID NO 44
 SOFTWARE: PatentIn Ver. 2.1
 LENGTH: 7
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: nucleotide
 OTHER INFORMATION: Codon binding sequence
 US-09-173-941-44

Query Match 87.9%; Score 29; DB 3; Length 7;
 Best Local Similarity 87.7%; Pred. No. 3e-05;
 Matches 6; Conservative 1; Mismatches 0;
 Indels 0; Gaps 0;

Qy 1 K\$ADLKR 7
 Db 1 K\$AEELKR 7

RESULT 8
 US-09-494-190-44
 Sequence 44, Application US/09494190
 GENERAL INFORMATION:
 APPLICANT: BARBAS, Carlos F.
 PATENT NO. 6610512
 TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
 FILE REFERENCE: TSRI 645.2
 CURRENT APPLICATION NUMBER: US/09/494,190
 CURRENT FILING DATE: 2000-01-28
 PRIOR APPLICATION NUMBER: EP/99/07742
 PRIOR FILING DATE: 1999-10-14
 PRIOR APPLICATION NUMBER: US 09/494,190
 PRIOR FILING DATE: 1998-10-16
 NUMBER OF SEQ ID NOS: 126
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 44
 LENGTH: 7
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: phage display
 OTHER INFORMATION: Selected and mutagenized
 US-09-494-190-44

Query Match Score 29; DB 4; Length 7;
Best Local Similarity 85.7%; Pred. No. 36+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 K\$ADLKR 7
Db 1 K\$ABLKR 7

RESULT 9
US-19-134-000C-6760
Sequence 6760, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROBACTER FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
CURRENT APPLICATION NUMBER: US/09/134,000C
FILE REFERENCE: 032796-032
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6760
LENGTH: 149
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-134-000C-6760

Query Match Score 29; DB 4; Length 149;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 K\$ADLKR 7
Db 46 K\$ADLKR 52

RESULT 10
US-09-540-236-3493
Sequence 3483, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATALINA
FILE REFERENCE: 2709.2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 3483
LENGTH: 251
TYPE: PRT
ORGANISM: M.catalzzhalis
US-09-540-236-3483

Query Match Score 29; DB 4; Length 251;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 K\$ADLKR 7
Db 186 K\$ADLKR 192

RESULT 11
US-08-485-355B-46
Sequence 46, Application US/08485355B
Patent No. 6170705
GENERAL INFORMATION:
APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in

Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fleur Hobbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treararin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE DOCUMENT NUMBER: A-38631-2/RFT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELELEX: 910 277299
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 73 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-08-485-355B-46
Query Match Score 28; DB 3; Length 73;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 SADLKR 7
Db 56 SADLKR 61

RESULT 12
US-09-489-039A-12780
Sequence 12780, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 09/489-039A
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 12780
LENGTH: 156
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12780
Query Match Score 28; DB 4; Length 156;
Best Local Similarity 71.4%; Pred. No. 60;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 K\$ADILKR 7
Db 46 K\$ADILKR 52

RESULT 13
US-09-543-681A-5089

; Sequence 509, Application US/09543681A
; Patent No. 6605703
GENERAL INFORMATION:
; APPLICANT: GRY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILI
; FILE REFERENCE: 2709_1002_001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5089
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Proteus mirabilis

US-09-543-681A-5089

Query Match 84.8%; Score 28; DB 4; Length 179;
Best Local Similarity 71.4%; Pred. No. 69;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 K\$ADILKR 7
Db 68 K\$ADILKR 74

RESULT 14
US-09-134-001C-4430

; Sequence 4430, Application US/09134001C
; Patent No. 6380370
GENERAL INFORMATION:
; APPLICANT: LYNN DOUCETTE-STAMM et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4430
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4430

Query Match 84.8%; Score 28; DB 4; Length 206;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 K\$ADILKR 7
Db 96 K\$KDLKR 102

RESULT 15
US-08-290-731C-11
; Sequence 11, Application US/08290731C
; Patent No. 5833646
; GENERAL INFORMATION:
; APPLICANT: BONTELL, David Douglas Lawrence

TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE SON OF SEVENLESS (mSOS) GENE,
TITLE OF INVENTION: AND mSOS POLYPEPTIDES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SUGHRAYE, MION, ZINN, MACPEAK & SEAS
STREET: 2100 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,731C
FILING DATE: 17-OCT-1994
CLASSIFICATION: 415
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU93/00068
FILING DATE: 17-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PLO921/92
FILING DATE: 17-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon
REGISTRATION NUMBER: 30,764
PRIORITY/DOCKET NUMBER: Q 36066
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7860
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-290-731C-11

Query Match 84.8%; Score 28; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 SADILKR 7
Db 71 SADILKR 76

Search completed: May 24, 2004, 14:53:23
Job time : 23 secs

GenCore version 5.1.6
right (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: May 24, 2004, 14:48:28 ; Search time 21 Seconds
(without alignments)
32.064 Million cell updates/sec

Title: US-09-807-665-41
Perfect score: 33
Sequence: 1 K\$ADLKR 7
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing First 45 summaries

Database : PIR_78:*

1: Pir1:
2: Pir2:
3: Pir3:
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	30	90.9	328	T37734	SUR4 family protein - fission yeast (<i>Schizosaccharomyces pombe</i>)
2	30	90.9	560	T48841	C;Species: <i>Schizosaccharomyces pombe</i>
3	30	90.9	1626	A75613	C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
4	29	87.9	648	T33339	C;Accession: T37734; T39101
5	29	87.9	1029	F87369	R;Wieder, H.; Duesterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G. submitted to the EMBL Data Library, September 1999
6	28	84.8	155	1 Q8CP1	A;Reference number: Z22742
7	28	84.8	155	G64140	A;Accession: T37734
8	28	84.8	155	2 B90713	A;Status: preliminary; translated from GB/EMBL/DDJB
9	28	84.8	155	F85563	A;Molecule type: DNA
10	28	84.8	155	2 AD0581	A;Cross-references: EMBL:AL117213; PIDN:CRB55289.1; GSPDB:GN00066; SPDB:SPAC1639.01C
11	28	84.8	156	2 A10317	R;Wieder, H.; Duesterhoeft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G. submitted to the EMBL Data Library, September 1999
12	28	84.8	180	2 B70202	A;Reference number: Z22827
13	28	84.8	183	2 S26961	A;Accession: T39101
14	28	84.8	321	2 T43455	A;Status: preliminary; translated from GB/EMBL/DDJB
15	28	84.8	348	1 S20911	A;Molecule type: DNA
16	28	84.8	477	2 G71319	A;Cross-references: EMBL:AL117212; PIDN:CAB55288.1; GSPDB:GN00066; SPDB:SPAC806.09C
17	28	84.8	763	2 S35908	R;Xiao, Y.P.; Thompson, D.K.; St Jean, A.; Charlebois, R.L.; Daniels, C.J. J. Bacteriol. 179, 6318-6324, 1997
18	28	84.8	803	2 T40475	C;Accession: T48841
19	28	84.8	818	2 T02436	N;Alternate names: chaperonin-containing Tcp-1
20	28	84.8	856	2 T31234	C;Species: <i>Halofexx volcanii</i>
21	28	84.8	988	2 T18986	C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 21-Jul-2000
22	28	84.8	1217	2 F69823	C;Accession: T48841
23	28	84.8	1297	2 S25714	R;Xiao, Y.P.; Thompson, D.K.; St Jean, A.; Charlebois, R.L.; Daniels, C.J. J. Bacteriol. 179, 6318-6324, 1997
24	28	84.8	1648	2 S71763	C;Title: Characterization of two heat shock genes from <i>Halofexx volcanii</i> : a model system
25	27	81.8	1274	1 HS2PB2	A;Reference number: Z24371; MUID:9744254; PMID:9355278
26	27	81.8	174	2 S59848	A;Accession: T48841
27	27	81.8	180	2 S56056	A;Status: preliminary; translated from GB/EMBL/DDJB
28	27	81.8	195	2 T36854	A;Cross-references: EMBL:AL117212; PIDN:CAB55288.1; GSPDB:GN00066; SPDB:SPAC806.09C
29	27	81.8	208	2 D87565	R;Xiao, Y.P.; Thompson, D.K.; St Jean, A.; Charlebois, R.L.; Daniels, C.J. J. Bacteriol. 179, 6318-6324, 1997

A;Molecule type: DNA
A;Residues: 1-560 <RIO>
C;Cross-references: EMBL:AF010470; PIDN:g2459985; PIDN:ABB81497.1; PIDN:g2459986
C;Genetics:
A;Gene: cct1
C;Superfamily: molecular chaperone t-complex-type
C;Keywords: heat shock

RESULT 3
A75613
hypothetical protein DRA0166 - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Ma, S.M.; Shen, M.; Vaithyanathan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75613
A;Accession: A75613
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1626 <WHI>
A;Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12309.1; PIDN:g646000
A;Experimental source: Strain R1
C;Genetics:
A;Gene: DRA0166
A;MNP position: 2
C;Superfamily: Deinococcus radiodurans hypothetical protein DRA0166

Query Match Score 30; DB 2; Length 560;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7
Db 315 RSSDLKR 321

RESULT 3
A75613
hypothetical protein DRA0166 - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Ma, S.M.; Shen, M.; Vaithyanathan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75613
A;Accession: A75613
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1626 <WHI>
A;Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12309.1; PIDN:g646000
A;Experimental source: Strain R1
C;Genetics:
A;Gene: DRA0166
A;MNP position: 2
C;Superfamily: Deinococcus radiodurans hypothetical protein DRA0166

Query Match Score 30; DB 2; Length 1626;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7
Db 522 KSADLRR 528

RESULT 4
T33319
hypothetical protein K07D4.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33319
R;Herkhau, J.; Wohldmann, P.
Submitted to the ENBL Data Library, July 1998
A;Description: The sequence of *C. elegans* cosmid K07D4.
A;Reference number: T21327
A;Accession: T33319
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-48 <HEN>
A;Cross-references: EMBL:AF077534; PIDN:AAC26284.1; GSPPDB:GN00020; CESP:K07D4.7
A;Experimental source: Strain Bristol N2; Clone K07D4
C;Genetics:
A;Gene: CESP:K07D4.7
A;Map position: 2
A;Indrcons: 23312; 517/3; 565/1; 606/3

Query Match Score 29; DB 2; Length 648;
Best Local Similarity 85.7%; Pred. No. 85;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KSADLKR 7
Db 45 KNADLKR 51

RESULT 5
F87369
TonB-dependent receptor [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: F87369
R;Nieman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.J.; Heidelberg, J.B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.N.; Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: F87369
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1039 <STO>
A;Cross-references: GB:AE005673; NID:g13422250; PIDN:AAK22954.1; GSPPDB:GN00148
A;Gene: CC0570

Query Match Similarity 87.9%; Score 29; DB 2; Length 1029;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7
Db 48 XSVDLKR 54

RESULT 6
OOCPI
Yba protein - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Accession: B24995; B64198
R;Asoh, S.; Matsuzawa, H.; Ishino, F.; Stroninger, J.L.; Matsushashi, M.; Ohta, T.; Eur. J. Biochem. 160, 231-238, 1986
A;Title: Nucleotide sequence of the pbpa gene and characteristics of the deduced amino acid sequence. A91176; MUID:87030266; PMID:353335
A;Accession: B24995
A;Molecule type: DNA
A;Residues: 1-155 <ASO>
A;Cross-references: GB:X04516; GB:D00001; GB:942313; PIDN:CAA28200.1; PIDN:G.R.; Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perria, N.T.; Burland, V.; Riley, M.; C. A.; Rose, D.J.; Maur, B.; Shao, Y.; Schenck, T.; 1453-1452, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; PMID:9278503
A;Accession: B64729
A;Molecule type: DNA
A;Residues: 1-155 <BLAT>
A;Cross-references: GB:AB00168; GB:U00096; NID:91786649; PIDN: AAC73737.1; PIDN:g1786855
A;Experimental source: strain K-12, substrate MG1655
C;Genetics:
A;Gene: yba
A;Map position: 15 min
A;Start codon: GTG
C;Superfamily: uncharacterized conserved protein HI0033

Query Match Similarity 84.8%; Score 28; DB 1; Length 155;
Best Local Similarity 71.4%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7
Db 45 KNADLKR 51

RESULT 7 G64140 YbeA protein homolog HI0033 - <i>Haemophilus influenzae</i> (strain Rd KW20) C;Species: <i>Haemophilus influenzae</i> C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 12-Jun-2003 C;Accession: G6140 R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Brandon, R.C.; Fine, L.D.; Scott, J.; Shirley, R.; Liu, L.; Kelley, A.; Glodek, A.; Weidman, D.M.; Goedayne, J.D.; McDonald, I.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A;Title: Whole-genome random sequencing and assembly of <i>Haemophilus influenzae</i> Rd. A;Reference number: A64000; MUID:95350630; PMID:7542800 A;Accession: G6140 A;Status: nucleic acid sequence not shown; translation not shown A;Molecule type: DNA A;Residues: 1-155 <PAR> A;Cross-references: GB:U32688; GB:L42023; NID:91572966; PIDN:AAC21711.1; PID:g1572978; T A;Start codon: GTG C;Genetics: A;Superfamily: uncharacterized conserved protein HI0033 Query Match Score 28; DB 2; Length 155; Best Local Similarity 71.4%; Pred. No. 34; Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0; Qy 1 KSADLKR 7 Db 45 KNADIKR 51	RESULT 10 AD0581 Conserved hypothetical protein STY0692 [imported] - <i>Salmonella enterica</i> subsp. <i>enterica</i> C;Species: <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi A;Note: this species has also been called <i>Salmonella typhi</i> C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 12-Jun-2003 C;Accession: AD0581 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, S.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.; Nature 413, 848-852, 2001 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K. A;Title: Complete genome sequence of a multiple drug resistant <i>Salmonella enterica</i> sero A;Reference number: AB0502; MUID:21534947; PMID:11677608 A;Accession: AD0581 A;Status: Preliminary A;Molecule type: DNA A;Residues: 1-155 <PAR> A;Cross-references: GB:AL513382; PIDN:CAD05118.1; PID:g16501892; GSPDB:GN00176 C;Genetics: A;Gene: STY0692 C;Superfamily: uncharacterized conserved protein HI0033 Query Match Score 28; DB 2; Length 155; Best Local Similarity 71.4%; Pred. No. 34; Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0; Qy 1 KSADLKR 7 Db 45 KNADIKR 51
RESULT 8 B90713 Hypothetical protein ECs0674 [imported] - <i>Escherichia coli</i> (strain O157:H7, substrain RI C;Species: <i>Escherichia coli</i> C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 12-Jun-2003 C;Accession: B90713 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Sasawata, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001 A;Title: Complete genome sequence of enterohemorrhagic <i>Escherichia coli</i> O157:H7 and gene A;Accession: B90713 A;Reference number: A99629; MUID:21156231; PMID:11258796 A;Status: Preliminary C;Genetics: A;Molecule type: DNA A;Residues: 1-155 <PAR> A;Cross-references: GB:BA000007; PIDN:BAB30097.1; PID:g13360132; GSPDB:GN00154 A;Experimental source: strain O157:H7, substrain RIMD 050952 C;Superfamily: uncharacterized conserved protein HI0033 Query Match Score 28; DB 2; Length 155; Best Local Similarity 71.4%; Pred. No. 34; Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0; Qy 1 KSADLKR 7 Db 45 KNADIKR 51	RESULT 11 A10317 Conserved hypothetical protein YPO2605 [imported] - <i>Yersinia pestis</i> (strain CO92) C;Species: <i>Yersinia pestis</i> C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 12-Jun-2003 C;Accession: A10317 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B. deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B.; Nature 413, 523-527, 2001 A;Title: Genome sequence of <i>Yersinia pestis</i> , the causative agent of plague. A;Reference number: AB0001; MUID:21470413; PMID:11586360 A;Accession: A10317 A;Status: Preliminary A;Molecule type: DNA A;Residues: 1-155 <PAR> A;Cross-references: GB:AL590842; PIDN:CA92848.1; PID:g15980592; GSPDB:GN00175 C;Genetics: A;Gene: YPO2605 C;Superfamily: uncharacterized conserved protein HI0033 A;Title: Genome sequence of enterohemorrhagic <i>Escherichia coli</i> O157:H7.
RESULT 9 F85563 Hypothetical protein ybaA [imported] - <i>Escherichia coli</i> (strain O157:H7, substrain EDL93 C;Species: <i>Escherichia coli</i> C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 12-Jun-2003 C;Accession: F85563 R;Perna, N.T.; Blunkett III, G.; Burland, V.; Muu, B.; Glasner, J.D.; Rose, D.J.; Mayhew, Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamitis, K.; Apodaca, Nature 409, 529-533, 2001 A;Title: Genome sequence of enterohemorrhagic <i>Escherichia coli</i> O157:H7.	RESULT 11 A10317 Conserved hypothetical protein YPO2605 [imported] - <i>Yersinia pestis</i> (strain CO92) C;Species: <i>Yersinia pestis</i> C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 12-Jun-2003 C;Accession: A10317 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B. deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B.; Nature 413, 523-527, 2001 A;Title: Genome sequence of <i>Yersinia pestis</i> , the causative agent of plague. A;Reference number: AB0001; MUID:21470413; PMID:11586360 A;Accession: A10317 A;Status: Preliminary A;Molecule type: DNA A;Residues: 1-155 <PAR> A;Cross-references: GB:AL590842; PIDN:CA92848.1; PID:g15980592; GSPDB:GN00175 C;Genetics: A;Gene: YPO2605 C;Superfamily: uncharacterized conserved protein HI0033 A;Title: Genome sequence of enterohemorrhagic <i>Escherichia coli</i> O157:H7.

Query Match Score 28; DB 2; Length 156;
 Best Local Similarity 71.4%; Pred. No. 35;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADIKR 7
 Db 45 KNADIKR 51

RESULT 12
 BT0202
 cytidylate kinase (cmk-2) homolog - Lyme disease spirochete
 C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: B70202
 R:Fraser, C.M.; Castens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, J.; Peterman, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, M.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Nature 390, 580-586, 1997.
 A:Authors: Smith, H.O.; Ventier, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.
 A:Accession: B70202
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Keywords: DNA
 A:Residues: 1-180 <KLE>
 A:Cross-references: GB:AE001180; GB:AB000783; NID:92688755; PIDN:ACC67164.1; PID:92688055
 A:Experimental source: strain B31
 C:Superfamily: *Methanococcus jannaschii* cytidylate kinase
 Query Match Score 28; DB 2; Length 180;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADIK 6
 Db 89 KSADIK 94

RESULT 13
 S26961
 H+-transporting two-sector ATPase (EC 3.6.3.1) delta chain - red alga (*Antithamnion* sp.).
 C:Species: chloroplast Antithamnion sp.
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 03-Jun-2002
 R:Kostrozwia, M.; Zetsche, K.
 J. Mol. Biol. 227, 961-970, 1992
 A:Reference number: S26957; PMID:1404401
 A:Accession: S26961
 A:Molecule type: DNA
 A:Residues: 1-183 <KOS>
 A:Cross-references: EMBL:X63382; NID:914170; PIDN:CAA44983.1; PID:g14176
 C:Genetics:
 A:Gene: atpD
 A:Genome: chloroplast
 C:SuperFamily: H+-transporting ATP synthase delta chain
 C:Keywords: ATP biosynthesis; chloroplast; hydrolase; thylakoid
 Query Match Score 28; DB 2; Length 183;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADIK 6
 Db 43 KSADIK 48

RESULT 14
 T43455
 hypothetical protein DKFZp434J1618.1 - human (fragment)
 C:Species: Homo sapiens (mar)
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000

Search completed: May 24, 2004, 14:52:48
 Job time : 22 secs

Result No.	Score	Query	Match	Length	DB	ID	Description
1	30	90.9	560	1	THS1_HALVO	030561	halobacteri
2	28	84.8	155	1	Y033_HAEIN	P44470	haemophilus
3	28	84.8	155	1	YBEA_BCOLY	Q8xf27	salmonella
4	28	84.8	155	1	YBEA_SALTY	Q9crj9	pasteurella
5	28	84.8	155	1	YJ23_PASMU	Q8dg3	yersinia pe
6	28	84.8	156	1	YQ05_YERPE	051759	borrelia bu
7	28	84.8	180	1	KCY2_BORBU	Q02849	antithiamino
8	28	84.8	183	1	ATPD_ANTSP	P49383	kluveromyco
9	28	84.8	348	1	ADH2_KLUFLA	Q83494	treponema p
10	28	84.8	477	1	Y481_TREPLA	Q06716	lactococcus
11	28	84.8	763	1	CLPL_LACLAA	Q9j6x9	homo sapien
12	28	84.8	970	1	K852_HUMCU	P51602	bacillus su
13	28	84.8	1217	1	YHR1_BACSU	Q02384	mus musculu
14	28	84.8	1232	1	SOS2_MOUSE	Q07890	homo sapien
15	28	84.8	1327	1	SOS2_HUMAN	P26239	homo sapien
16	28	84.8	1608	1	RRLP_BDV	P47171	stachyromyc
17	28	84.8	1648	1	YJ9H_YEAST	P04913	schizosacch
18	27	81.8	174	1	H2B1_SCHPO	P47913	schizosacch
19	27	81.8	174	1	RL10_YEAST	P58850	methanoscrc
20	27	81.8	297	1	YK46_METTAC	Q55328	treponema p
21	27	81.8	343	1	TP38_TREPA	Q9ryd0	thermotoga
22	27	81.8	353	1	PROB_THEMEA	Q971t1	sulfolobus
23	27	81.8	387	1	TP6A_SULTHO	Q05208	sulfolobus
24	27	81.8	389	1	TP6A_SULSH	Q92010	gallus galli
25	27	81.8	448	1	ZFG1_CHICK	Q43829	homo sapien
26	27	81.8	449	1	ZFG1_HUMAN	Q08376	mus musculu
27	27	81.8	449	1	ZFG1_MOUSE	Q84923	buchnera ap
28	27	81.8	482	1	EX1_BUCAPI	Q9gbq7	mycoplasma
29	27	81.8	513	1	ATPA_MYCIPU	P44225	haemophilus
30	27	81.8	520	1	VG29_HAEIN	Q9jg23	arabidopsis
31	27	81.8	522	1	C5P3_ARATH	Q5046	rhodospiril
32	27	81.8	1191	1	NIFJ_RHORU	Q09853	schizosacch
33	27	81.8	1616	1	YADE_SCHPO		

ALIGNMENTS						
RESULT 1						
THS1_HALVO	STANDARD;	PRT;	560 AA.			
ID THS1_HALVO						
AC O30561;						
DT 15-JUL-1998 (Rel. 36, Created)						
DT 15-JUL-1998 (Rel. 36, Last sequence update)						
DT 16-OCT-2001 (Rel. 40, Last annotation update)						
DB Thermosine subunit 1 (Heat shock protein CCT1).						
GN CCT1.						
OS Halobacterium volcanii (Halofexax volvani).						
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;						
OC Halobacteace; Halofexax.						
OC NCBI TaxID:2246;						
[1]						
RN						
SEQUENCE FROM N.A.						
RX MEDLINE=7474254; PubMed=9335278;						
RA Kuo Y.-P., Thompson D.K., St. Jean A., Charlebois R.L., Daniels C.J.;						
RT "Characterization of two heat shock genes from Halofexax volvani: a model system for transcription regulation in the Archaea."						
RL J. Bacteriol. 179:6318-6324 (1997).						
CC -I- FUNCTION: MOLECULAR CHAPERONE; BINDS UNFOLDED POLYPEPTIDES IN VITRO, AND HAS A WEAK ATPASE ACTIVITY (BY SIMILARITY).						
CC -I- SUBUNIT: FORMS AN OLIGOMERIC COMPLEX OF EIGHT-MEMBERED RINGS (BY SIMILARITY).						
CC -I- INDUCTION: By heat shock.						
CC -I- SIMILARITY: Belongs to the TCP-1 chaperonin family.						
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).						
CC PIR: T48841; T48841; -.						
DR PIR: AF010470; AB81497.1; -.						
DR PIR: T48841.						
DR InterPro: IP04424; Chaperonin TCP-1.						
DR InterPro: IP002194; Chaperonin CPn60.						
DR InterPro: IP001844; Chaperonin CPn60.						
DR InterPro: IP002433; CPn60/TCP-1.						
DR InterPro: IP008939; GroEL-ATPase.						
DR Pfam: PF001118; CPn60 TCP1; 1.						
DR PRINTS; PR00298; CHAPERONIN60.						
DR PRINTS; PR00304; TCOMPLEXTCP1.						
DR PROSITE; PS00750; TCP1_1; 1.						
DR PROSITE; PS00751; TCP1_2; 1.						
DR PROSITE; PS00995; TCP1_3; 1.						
DR Chaperone; ATP-Binding; Heat shock.						
DR SEQUENCE: 560 AA; 58325 MW; 88B73B2AD70DC341 CRC64;						
Query Match Score 90.9%; DB 1; Length 560;						
Best Local Similarity 85.7%; Pred. No. 21;						
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;						
QY 1 KSADLKR 7						

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete Genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and Genomic comparison with a laboratory strain K-12.";
 RT DNA Res. 8:11-22 (2001).
 RL [6]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.E. flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINES=22272406; PubMed=1238590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xie Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome Sequence of *Shigella flexneri* 2a: insights into pathogenicity
 through comparison with genomes of *Escherichia coli* K12 and O157.";
 RT Nucleic Acids Res. 30:4432-4441 (2002).
 RL [9]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.E. flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
 RX MEDLINES=2350274; PubMed=1270152;
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 RA Meau B., Perna N.T., Payne S.M., Ruryan-Janesky L.J., Zhou S.,
 RA Schwartz D.C., Blattner F.R.;
 RT "Complete genome sequence and comparative genomics of *Shigella*
 flexneri serotype 2a strain 2457T.";
 RL Infect. Immun. 71:2775-2786 (2003).
 RL [10]
 RP X-RAY CRYSTALLOGRAPHY (1.68 ANGSTROMS)
 RA Banach J., Shen J., Rost B., Xiao R., Acton T., Montelione G.,
 RA Hart J.F.;
 RT "Structure of *E. coli*;"
 CC -!- SIMILARITY: Belongs to the UPR0247 family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AB000168; CAA28200.1; -.
 DR EMBL: UB2598; AAC73737.1; -.
 DR EMBL: D90704; BA35233.1; -.
 DR EMBL: AE016757; AAN73200.1; -.
 DR EMBL: AE05242; AAG5197.1; -.
 DR EMBL: AE015055; AAN42281.1; -.
 DR EMBL: AE016980; AAP16152.1; -.
 DR PIR: B24995; QCEP1.
 DR PIR: B30713; B90713.
 DR PIR: B85563; R85563.
 DR PDB; IN55; 04-FBB-03.
 DR EcoGene; EG11254; YbeA.
 DR HAMAP; MF_00658; " 1.
 DR InterPro; IPR003742; DUF163.
 DR Pfam; PF02590; DUF163.
 DR Hypothetical protein; Complete proteome; 3D-structure.
 PT CONFLICT 59 59 Q -> L (IN REF. 9).
 SQ SEQUENCE 155 AA; 17341 MW; DFA05FCF4EB057E7 CRC64;
 Query Match 84.8%; Score 28; DB 1; Length 155;
 Best Local Similarity 71.4%; Pred. No. 16;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 45 KNADIKR 51

RESULT 4
 YBEA SALTY STANDARD PRT; 155 AA.
 ID AC Q8XP77;
 ID DT 10-OCT-2003 (Rel. 42, Created)
 ID DT 10-OCT-2003 (Rel. 42, Last annotation update)
 ID DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DB Hypothetical UPF0247 protein in YbeA
 GN YBEA OR STM061 OR STM0692 OR T2226.
 OS *Salmonella typhimurium*, and
Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Salmonella*.
 OC NCBI_TaxID=602, 601; .
 RN NCBI_TaxID=602, 601; .
 RN [1]
 RN SEQUENCE FROM N.A.
 RP SPECIES=S.typhimurium; STRAIN=LIT2 / SGSC1412 / ATCC 700720;
 RC SPECIES=S.typhimurium; STRAIN=CT18 / ATCC 700931;
 RX MEDLINE=21534348; PubMed=11677659;
 RA McClelland M., Sanderson K.B., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwoll S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvey E.,
 RA Ryan E., Sun H., Plorea L., Miller W., Stoecking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar *Typhimurium*
 LIT2.";
 RT RT
 RT RT
 RN Nature 413:852-856 (2001).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP SPECIES=S.typhi; STRAIN=CT18;
 RC SPECIES=S.typhi; STRAIN=CT18;
 RX MEDLINE=21534347; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moulé S., O'Gaora P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skeatton J., Stevens K.,
 RA Whitehead S., Barrell B.G.;
 RT "Complete genome sequence of multiple drug resistant *Salmonella*
 enterica serovar *Typhi* CT18.";
 RT RT
 RN Nature 413:848-852 (2001).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;
 RC SPECIES=S.typhi; STRAIN=TY2;
 RX MEDLINE=22531167; PubMed=12444504;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D.J.,
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyann V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of *Salmonella enterica* serovar *Typhi* strains Ty2
 and CT18.";
 RT RT
 RL J. Bacteriol. 185:2330-2337 (2003).
 CC -!- SIMILARITY: Belongs to the UPR0247 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE008725; AAU19592.1; -.
 CC EMBL: AL627257; CAD05118.1; -.
 CC EMBL: AE016841; AAC069829.1; -.
 CC StyGene: SG??????; YbeA.
 CC HAMAP; MF_00658; " 1.
 CC InterPro; IPR003742; DUF163.
 CC DR Pfam; PF02590; DUF163.
 CC DR TIGR00246; TIGR00246; 1.
 CC DR Hypothetical protein; Complete proteome; 3D-structure.
 CC KW SEQUENCE 155 AA; 17199 MW; D34EFDF4EB9D36A CRC64;
 CC SQ

1 KSAIDLKR 7
 | :| :| :|

Query Match	84.8%	Score 28;	DB 1;	Length 155;	RX MEDLINE=21470413; PubMed=11583360;
Best Local Similarity	71.4%	Pred. NO. 16;			RA Parkhill J.W., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
Matches	2;	Mismatches	0;	Indels	RA Prentice M.B., Sebaiia M., James K.D., Churcher C., Mungall K.L.,
Conservative					RA Baker S., Basham D., Bentley S.D., Brooks K., Cedeno-Tarraga A.M.,
Qy	1 KSADIKR 7				RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
Db	45 KNADIKR 51				RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
					RA Leather S., Moule S., Cyson P.C.F., Quail M.A., Rutherdale K.,
					RA Simmonds M., Skilton J., Stevens K., Whitehead S., Barrell B.G.,
					RT "Genome sequence of <i>Yersinia pestis</i> , the causative agent of plague.";
RESULT 5					RL Nature 413:523-527 (2001).
YJ23_PASMU		STANDARD;	PRT;	155 AA.	RN [2]
ID YJ23_PASMU					RP SEQUENCE FROM N.A.
AC Q9CUE9					RC STRAIN/KIMS / Biovar Mediaeavalis;
DT 10-OCT-2003 (Rel. 42, Created)					RX MEDLINE=22137863; PubMed=12142320;
DT 10-OCT-2003 (Rel. 42, Last sequence update)					RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
DT 10-OCT-2003 (Rel. 42, Last annotation update)					RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
DE Hypothetical UPF0247, protein PM1923.					RA Fetherston J.D., Lindler L.E., Brubaker R.R., Piano G.V.,
GN PM1923.					RA Straley S.C., McElroy K.A., Nilles M.L., Matson J.S., Blattner F.R.,
OS <i>Pasteurella multocida</i> .					RA Perry R.D.,
OC <i>Bacteriaceae</i> ; <i>Proteobacteria</i> ; <i>Gammaproteobacteria</i> ; <i>Pasteurellales</i> ;					RT "Genome sequence of <i>Yersinia pestis</i> KIM.";
OC <i>Pasteurellaceae</i> ; <i>Pasteurellales</i> .					RL Bacteriol. 184:460-461 (2002).
OX NCBI_TaxId=747;					CC - SIMILARITY: Belongs to the UPF0247 family.
RN [1]					CC
SEQUENCE FROM N.A.					This SWISS-PROT entry is copyright. It is produced through a collaboration
RC STRAIN=PM010;					CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RX MEDLINE=21145866; PubMed=11248100;					CC the European Bioinformatics Institute. There are no restrictions on its
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;					use by non-profit institutions as long as its content is in no way
RT "Complete genomic sequence of <i>Pasteurella multocida</i> PM70.";					modified and this statement is not removed. Usage by and for commercial
RL Proc. Natl. Acad. Sci. U.S.A. 98:3455-3455 (2001).					entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC - SIMILARITY: Belongs to the UPF0247 family.					CC
CC SEQUENCE FROM N.A.					DR AJ414153; AACM92848.1;
RC STRAIN=PM010;					DR EMBL; AB013721; AACM934756.1;
CC MEDLINE=21145866; PubMed=11248100;					DR PIR; AI0317; AI0317.
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;					DR HAMAP; MF_00658;
RT "Complete genomic sequence of <i>Pasteurella multocida</i> PM70.";					DR IPR0001742;
RL Proc. Natl. Acad. Sci. U.S.A. 98:3455-3455 (2001).					DR InterPro; IPR0001742;
CC - SIMILARITY: Belongs to the UPF0247 family.					CC pFam; PF02580; DUP163;
CC SEQUENCE FROM N.A.					CC TIGRFAMS; TIGR00246; TIGR00246; 1.
DR EMBL; AE06229; AAK04007.1; -.					CC Hypothetical protein; Complete proteome.
DR HAMAP; MF_00658;					KW KW
DR InterPro; IPR003742; DUP163.					SEQUENCE 156 AA; 17520 MW; 612248883BFEC2 CRC64;
DR PF0163; DUP163; 1.					SQ SQ
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					RESULT 7
DR Hypothetical protein; Complete proteome.					KC72_BORBU STANDARD; PRT; 180 AA.
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					ID KC72_BORBU
DR EMBL; AE06229; AAK04007.1; -.					AC 0S1159;
DR HAMAP; MF_00658;					DT 16-OCT-2001 (Rel. 40, Created)
DR InterPro; IPR003742; DUP163.					DT 16-OCT-2001 (Rel. 40, Last sequence update)
DR PF0163; DUP163; 1.					DT 28-FEB-2003 (Rel. 41, Last annotation update)
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					DE Cytidine kinase 2 (EC 2.7.4.14) (CK 2) (Cytidine monophosphate
DR Hypothetical protein; Complete proteome.					DE kinase 2) (CMP kinase 2).
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					GN CMK2 OR BB0819.
DR EMBL; AE06229; AAK04007.1; -.					OS <i>Borrelia burgdorferi</i> (Lyme disease spirochete).
DR HAMAP; MF_00658;					OX NCBI_TaxID=139;
DR InterPro; IPR003742; DUP163.					RN SEQUENCE FROM N.A.
DR PF0163; DUP163; 1.					RC STRAIN=TATCC 35210 / B31;
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					RA MEDLINE=98065943; PubMed=9403685;
DR Hypothetical protein; Complete proteome.					RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					RA Lathigra R., White O., Dodson K.A., Hickey E.K., Gwinn M.,
DR EMBL; AE06229; AAK04007.1; -.					RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D., Hanson M.,
DR HAMAP; MF_00658;					RA Peterson J., Karlaftis A.R., Quackenbush J., Salzberg S., Hanson M.,
DR InterPro; IPR003742; DUP163.					RA van Vugt R., Palmer N., Adams M.D., Goedeyne J.D., Weidman J.,
DR PF0163; DUP163; 1.					RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					RC STRAIN=CO-92 / Biovar Orientalis;
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMBL; AE06229; AAK04007.1; -.					
DR HAMAP; MF_00658;					
DR InterPro; IPR003742; DUP163.					
DR PF0163; DUP163; 1.					
DR TIGRFAMS; TIGR00246; TIGR00246; 1.					
DR Hypothetical protein; Complete proteome.					
SEQUENCE 155 AA; 17440 MW; 0A7E9B83D9435A7 CRC64;					
DR EMB					

RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.";
 RT NATURE 390:560-566 (1997).
 RL CC -!- CATALYTIC ACTIVITY: ATP + (d)CDP + (d)CDP.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the cytidylate kinase family. Subfamily 2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AB001180; AAC67164.1; -.
 DR PIR; B70202; B70202.
 DR TIGCR; BB0819; -!
 DR HANAP; MF_00239; -!
 DR Transferase; Kinase; ATP-binding; Complete proteome.
 KW FT NP BIND 7 15 ATP (POTENTIAL)
 SEQUENCE NP BIND 7 15 ATP (POTENTIAL)
 Qy Query Match 84.8%; Score 28; DB 1; Length 180;
 Best Local Similarity 100.0%; Fred. No. 18;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 SEQUENCE 180 AA; 21066 MW; 4CD956B6F8A590B CRC64;
 DR Qy 1 KSADIK 6
 Db 89 KSADIK 94
 RESULT 8
 ATPD_ANTSP STANDARD; PRT; 183 AA.
 ID ATPD_ANTSP STANDARD; PRT; 183 AA.
 AC Q02849;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP synthase delta chain, chloroplast (EC 3.6.3.14).
 GN ATPD.
 OS Autithamnion sp.
 OG Chloroplast.
 OC Eukaryota; Rhodophytta; Florideophyceae; Ceramiales; Ceramiaceae;
 OC Autithamnion.
 RN NCBI_TaxID:2767;
 RP SEQUENCE FROM N.A.
 RC STRAIN=LB 95-79;
 RX MEDLINE=93021132; PubMed=1404401;
 RA Kostrzewa M., Zetsche K.;
 RT Large ATP synthase operon of the red alga *Autithamnion* sp. resembles
 RT the corresponding operon in cyanobacteria.";
 RL J. Mol. Biol. 227: 961-970 (1992).
 CC -!- FUNCTION: This protein seems to be part of the stalk that links
 CC CF(0) to CF(1). It either transmits conformational changes from
 CC CF(0) into CF(1) or is implicated in proton conduction.
 CC -!- CATALYTIC ACTIVITY: ATP + H⁽²⁾O + H(+)(In) = ADP + phosphate +
 CC H(+)(Out).
 CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
 CC core - and CF(0) - the membrane proton channel. CF(1) has five
 CC subunits: alpha(1), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
 CC has three main subunits: a, b and c.
 CC -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
 CC -!- SIMILARITY: Belongs to the ATPase delta chain family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; X63382; CAA4983.1; -.
 DR PIR; S26961; S26961.
 DR InterPro; IPR00711; ATPsynth_OSCP.
 DR Pfam; PF0213; OSCP; 1.
 DR PRINTS; PR0012; ATPaseBLTA.
 DR TIGCR; TIGR01145; ATP synthetase delta; 1.
 DR PROSITE; PS00389; ATPase_DELTA; 1.
 KW Hydrolase; ATP synthase; CF(1); Chloroplast; Hydrogen ion transport.
 SQ SEQUENCE 183 AA; 20601 MW; 51C04FP66218AEBA CRC34;
 Query Match 84.8%; Score 28; DB 1; Length 183;
 Best Local Similarity 100.0%; Fred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DR Qy 1 KSADIK 6
 Db 43 KSADIK 48
 RESULT 9
 ADH2_KLUTA STANDARD; PRT; 348 AA.
 ID ADH2_KLUTA STANDARD; PRT; 348 AA.
 AC P49380;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Alcohol dehydrogenase II (EC 1.1.1.1).
 GN ADH2.
 OS Kluyveromyces lactis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae;
 OC NCBI_TaxID=28985;
 RN RP SEQUENCE FROM N.A.
 RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;
 RX MEDLINE=92269769; PubMed=1568917;
 RA Chain D.H., Salvador C., Denis C.J.;
 RA "Evolution of the alcohol dehydrogenase (ADH) genes in yeast:
 RT characterization of a fourth ADH in Kluyveromyces lactis.";
 RL Mol. Gen. Genet. 232:479-488 (1992).
 DR -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
 CC NADH.
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -!- SUBUNIT: Homotrimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cyttoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
 CC family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; X63397; CAA45739.1; -.
 DR PIR; S20911; S20911.
 DR InterPro; IPR002328; ADH_zinc.
 DR Pfam; PF0107; ADH_zinc_N; 1.
 DR PROSITE; PS00059; ADH_ZINC; 1.
 KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
 FT METAL 44 44 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 FT METAL 67 67 ZINC 2 (BY SIMILARITY).
 DR METAL 98 98 ZINC 3 (BY SIMILARITY).
 PT METAL 101 101 ZINC 4 (BY SIMILARITY).
 PT METAL 104 104 ZINC 5 (BY SIMILARITY).
 PT METAL 112 112 ZINC 6 (BY SIMILARITY).
 PT METAL 154 154 ZINC 1 (CATALYTIC) (BY SIMILARITY).

SQ SEQUENCE 348 AA; 37097 MW; F3B64AE1FF520689C CRC64;
 Query Match 84.8%; Score 28; DB 1; Length 348;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLK 6
 Db 164 KSADLK 169

RESULT 10
 Y481_TREPA STANDARD PRT; 477 AA.
 ID Y481_TREPA SEQUENCE FROM N.A.
 AC 083194;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Hypothetical protein TP0481.
 GN TP0481.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_TaxId=160;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RX MEDLINE=9665876; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
 RA Dodson R., Gwynn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardin J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of *Treponema pallidum*, the syphilis
 RT spirochete";
 RJ Science 281:375-388(1998).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions. There are no restrictions on its
 CC modification and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; AE001224; AAC65470.1; -.
 DR TIGR; TP0481; -.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 107 129 POTENTIAL.
 SQ SEQUENCE 477 AA; 53153 MW; D32B041577BE4120 CRC64;
 Query Match 84.8%; Score 28; DB 1; Length 477;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 SADLR 7
 Db 87 SADLR 92

RESULT 11
 CLPL_LACLA STANDARD PRT; 763 AA.
 ID CLPL_LACLA SEQUENCE FROM N.A.
 AC Q06716;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DE ATP-dependent protease ATP-binding subunit clpL.
 GN CLPL.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Plasmid pUC22.

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
 OC NCBI_TaxId=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CNPB2_270; PubMed=8387149;
 RX MEDLINE=9334749; Huang D.C., Huang X.F., Novel G., Novel M.;
 RA "Two genes present on a transposon-like structure in *Lactococcus* lactis are involved in a Clp family proteolytic activity.";
 RT Mol. Microbiol. 7:957-965(1993).
 RL FUNCTION: COULD BE THE ATP-DEPENDENT SPECIFICITY COMPONENT OF AN
 CC ATP-DEPENDENT PROTEASE.
 CC |- SIMILARITY: Belongs to the clpA/clpB family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC DR EMBL; X62333; CBA44207.1; -.
 DR PIR; S35908; S35908.
 DR InterPro; IPR001593; AAA_ATPase.
 DR InterPro; IPR003950; AAA_ATPase_centr.
 DR InterPro; IPR001270; Chapermin_cTPA/B.
 PFam; PF00004; AAA; 1.
 PRINTS; PRO0300; CLPROTEASEA.
 SMART; SM00382; AAA; 2.
 DR PROSITE; PS000870; CLPAB_1; FALSE NEG.
 DR PROSITE; PS000871; CLPAB_2; FALSE NEG.
 KW Chaperone; ATP-binding; Repeat; Plasmid.
 DOMAIN 136 386 I.
 FT DOMAIN 437 629 II.
 PT NR_BIND 181 188 ATP (POTENTIAL).
 PT NP_BIND 511 518 ATP (POTENTIAL).
 SQ SEQUENCE 763 AA; 86018 MW; E92CCF2E629D9C14 CRC64;
 Query Match 84.8%; Score 28; DB 1; Length 763;
 Best Local Similarity 71.4%; Pred. No. 92;
 Matches 5; Conservative 4; 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSAADLKR 7
 Db 411 KAADIKR 417

RESULT 12
 KB2_HUMAN STANDARD PRT; 970 AA.
 ID KB2_HUMAN STANDARD;
 AC Q9Y6X9; Q9UF28; Q9YCV2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protein_KIAA0852.
 GN Homo sapiens (Human).
 OS Homo sapiens (Human). Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=99156230; PubMed=10048485;
 RA Nagase T., Ishikawa K.-I., Suayama M., Kikuno R., Hiroseawa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.",
 RT DNA Res. 5:355-364 (1998).
 RL [12]
 RP SEQUENCE FROM N.A.

RA	Lee S.M., Levine A., Liu H., Masuda S., Muel C., Nedigre C., Medina N., Meiliado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.N., Portelle D., Porwollik S., Prescott A.M., Prosecan E., Pujic J., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadde V., Sato T., Scanlan B., Schleich S., Schroeter R., Scuffone F., Seixas J., Sekowska A., Seror S.J., Sevor P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Togoni A., Tomoto V., Uchiyama S., Vandembrou M., Vanier A., Wambt R., Wedler E., Weltzien H., Wetzger T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshioka H.F., Zumstein B., Yoshikawa H., Danchin A.; "The complete genome sequence of the Gram-positive bacterium <i>Bacillus subtilis</i> ", <i>Nature</i> 390:249-256 (1997).
CC	-!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.
CC	-!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce or send an email to license@isb-sib.ch).
DR	ENBML: X96983; CAB465702; 1; -. PFR: F9932; F69823. PFR: P00644; 182P. DR SubtilList: BG11596; yhcr. InterPro: IPR008334; 5'-Nucleotidase_C. InterPro: IPR006116; 5'-Nucleotidase_N. InterPro: IPR006179; 5_nucleotidase_. InterPro: IPR01899; C7m_pos_anchor. InterPro: IPR004843; N-peptidase. InterPro: IPR006021; SNase. InterPro: IPR004165; tRNA_anti. Pfam: PF02872; 5_nucleotideC_1. DR Pfam: PF00746; Gram_pos_anchor; 1. DR Pfam: PF00149; MetalLiphos; 1. DR Pfam: PF00565; Snase; 1. DR Pfam: PF01333; tRNA; anti; 2. DR Pfam: PF01607; APYRASE_FAMILY. SMART: SM00318; SNC; 1. PROSITE: PS00785; 5' NUCLEOTIDASE_1; FALSE_NEG. PROSITE: PS00786; 5' NUCLEOTIDASE_2; 1. PROSITE: PS00830; TNASE; 3; 1. Hypothetical protein_Hydrolase; Signal; Complete proteome.
FT	SIGNAL 1 46 CHAIN 47 1217 POTENTIAL PROTEIN YHCR. DOMAIN 376 517 TRASe-LIKE. SEQUENCE 1217 AA; 132686 MW; 7973160CC898DC0B CRC64;
SQ	Query Match Similarity 84.8%; Score 28; DB 1; Length 1217; Best Local Similarity 100.0%; Pred. No. 1e+02; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Dy	1 KADLKK
Db	358 KADLKK 363

RESULT 15
 SOS2_HUMAN STANDARD; PRT; 1332 AA.
 ID Q1890; Q15503; 38. Created)
 DT 15-JUL-1999 (Rel. 38. Last sequence update)
 DT 10-OCT-2003 (Rel. 42. Last annotation update)
 Son of sevenless protein homolog 2 (SOS-2).
 SOS2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 OX [1]_TaxID=3606;
 RN SEQUENCE FROM N.A.
 RP TISSUE:Brain;
 RX PMID=93262494; PubMed=8493519;
 RA Chardin P., Camonis J.H., Gale N.W., van Aelst L., Wigler M.H.,
 RA "Human Sos1: a guanine nucleotide exchange factor for Ras that binds
 RT to GRB2.";
 RL Science 260:1338-1343(1993).
 RN [2]
 RP SEQUENCE OF 724-1296 FROM N.A.
 RA TISSUE:Placenta;
 RA Path I., Apion F., Dutrillaux B., Tocque B.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -- FUNCTION: Promotes the exchange of Ras-bound GDP by GTP (By
 CC similarity). Contains 1 DBU-homology (DH) domain.
 CC -- SIMILARITY: Contains 1 Ras-GEF domain.
 CC -- SIMILARITY: Contains 1 PH domain.
 CC -- SIMILARITY: Contains 1 RhoGEF domain.
 CC -- SIMILARITY: Contains 1 RAS-GEF domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; L13888; AAA3594.1; .
 DR EMBL; L20686; AAA9182.1; .
 DR IBSPI; Q62245; IPMS.
 DR Genew; HGNC:11188; SOS2.
 DR MIM; 601247; .
 DR GO; GO:0005085; F:guanyl-nucleotide exchange factor activity; NAS.
 DR GO; GO:0007264; P:small GTPase mediated signal transduction; NAS.
 DR InterPro; IPR01331; CDS_CDC24.
 DR InterPro; IPR00849; PH.
 DR InterPro; IPR008937; Ras_GEF.
 DR InterPro; IPR000651; RasGPN.
 DR InterPro; IPR001895; RasGFR_CD025.
 DR InterPro; IPR000219; RhoGEF.
 DR PFAM; PF00069; PH; 1.
 DR PFAM; PF00617; RasGEP; 1.
 DR PFAM; PF00618; RasGEP; 1.
 DR PROSITE; PS00521; RhoGEP; 1.
 DR SMART; SM00333; PH; 1.
 DR SMART; SM00147; RasGEP; 1.
 DR SMART; SM00229; RasGEP; 1.
 DR PROSITE; PS01074; DH; 1; FALSE_NEG.
 DR PROSITE; PS00720; RhoGEP; 1.
 DR PROSITE; PS00003; PH_DOMAIN; 1.
 DR PROSITE; PS50009; RAGEF_CAT; 1.
 DR PROSITE; PS50212; RAGEF_NTER; 1.
 KW Guanine-nucleotide releasing factor.
 FT DOMAIN 198 388 DH.
 FT DOMAIN 442 546 PH.

FT DOMAIN 595 739 N-TERMINAL_RAS-GEF.
 FT DOMAIN 778 1017 RAS_GEF.
 FT DOMAIN 755 758 POLY-PRO.
 FT DOMAIN 1183 1208 POLY-PRO.
 FT DOMAIN 1203 1208 POLY-PRO.
 FT CONFLICT 778 778 D -> H (IN REF. 2).
 FT CONFLICT 861 861 Q -> R (IN REF. 2).
 FT CONFLICT 948 948 R -> K (IN REF. 2).
 FT CONFLICT 999 999 S -> C (IN REF. 2).
 FT CONFLICT 1032 1032 S -> E (IN REF. 2).
 FT CONFLICT 1042 1042 T -> A (IN REF. 2).
 FT CONFLICT 1112 1114 CGS -> LC (IN REF. 2).
 FT CONFLICT 1159 1159 A -> G (IN REF. 2).
 FT CONFLICT 1296 1296 Q -> B (IN REF. 2).
 SQ SEQUENCE 153030 MW; EB4BA9E52FA85AC9 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 1332;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 SADLR 7
 Db 669 SADLR 674

Search completed: May 24, 2004, 14:51:22
 Job time : 13 secs

No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Abrial J.F., Doyle C., Baxter E.G., Holt G., Nelson C.R., Miklos G.L.G., DR InterPro; IPR071101; 19-like. RA InterPro; IPR003538; 19_C2. RA InterPro; IPR003006; 19_MHC. RA InterPro; IPR0047; 19_14. RA Pfam; PF0047; 19_14. RA SMART; SMD0418; IgC2; 16. RA PROSITE; PS50835; Ig_LIKE; 14. RA PROSITE; PS00290; Ig_MHC; 1. RA Immunoglobulin domain.

RA de Pablos B., Delcher A., Deng I., Dew I., Dietz S.M., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA DR GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Dodson K., Doup L.E., Downes M., Dugat-Rocha S., Dunkov B.C., Dunn P., DR InterPro; IPR071101; 19-like. RA InterPro; IPR003538; 19_C2. RA InterPro; IPR003006; 19_MHC. RA InterPro; IPR0047; 19_14. RA Pfam; PF0047; 19_14. RA SMART; SMD0418; IgC2; 16. RA PROSITE; PS50835; Ig_LIKE; 14. RA PROSITE; PS00290; Ig_MHC; 1. RA Immunoglobulin domain.

RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Hostin D., Houston T.K.A., Howland T.J., Wei M.-H., Bergmann C., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Jalai M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Lai Z., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Lasko P., Lei Y., Levittky A.A., Li J., Li Z., Liang Y., Lin X., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Mukulov G., Milshina N.V., Mobarry C., Morris J., Moskrefi A., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Mount S.M., Moy M., Murphy L., Murphy D.M., Nelson D.L., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclob J.M., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Palazzolo M., Pettmann G.S., Pan S., Puris V., Reese M.G., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Shue B.C., Siden-Kiamos I., Simpson M.P., Smith T., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Svartas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Wang Z.-Y., Wasserman D.A., Weintraub G.M., Weissenbach J., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Williams S.M., Zawley R.F., Worley K.C., Wu D., Yang S., Yao Q.A., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C., Rubin G.M., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RT "The Genome sequence of *Drosophila melanogaster*," DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RL Science 287:2185-2195 (2000). DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RN [12] SEQUENCE FROM N.A. RP

RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Carlson J.W., Center A., Champé M., Davenport B., Diez S.M., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Carlson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Ferriera S., Frise E., Gallo R.F., Garg N.S., George R.A., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Ikegami C., Jalali M., Kruse D., Li P., Mattei B., Moskrefi A., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Mcintosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuncio J., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Paclob J., Paragas V., Park S., Patel S., Pfeiffer B., Richards S., Scheeler F., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Phouanenavong S., Pittman G.S., Puris V., Scheeler F., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RT "Sequencing of *Drosophila melanogaster* genome," DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases. DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RN [13] SEQUENCE FROM N.A. RP

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Hrdlicky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Tupay J.L., Berman C., Carlson J.W., Celniker S.E., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Clamp M., Drysdale R., Emmert D., Frieze E., de Grey A., Harris N., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield B., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RT "Annotation of *Drosophila melanogaster* genome," DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases. DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RN [14] SEQUENCE FROM N.A. RP

RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases. DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RN [15] SEQUENCE FROM N.A. RP

RA FlyBase; DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases. DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

CC "!! SIMILARITY: CONTAINS 15 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS. DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA FlyBase; DR GO; GO:0004685; F: calcium/calmodulin-dependent protein kinase. . . ; IDA. RA GO; GO:004687; F: myosin-light-chain kinase activity; IDA.

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*,";
 RL *Science* 287:2185-2195 (2000).
 [2]
 RN
 SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzen J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A.W., Chapple M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup J., Dresneak D., Farfan D.,
 RA Ferreira J., Frise E., Galle R.F., Garg N.J., George R.A.,
 RA Gonzalez M., Houck C., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibeagha-Ajadi M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunno J.,
 RA Pacioli J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phanomenang S., Pittman G.S., Puriv V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svistkis R., Tector C., Tyler D.,
 RA Williams S.M., Zavest J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of *Drosophila melanogaster* genome,";
 RL Submitted (MAR 2000) to the ENSEMBL/genBank/DBJ databases.

RN [3] SEQUENCE FROM N.A.
 RP Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Huang Y., Xaminder S.E., Prochnik S.E., Smith C.D.,
 RA Hradecky P., Tupy J.L., Bergman C., Carlson J.W., Celtniker S.B.,
 RA Clamp M., Drysdale R., Ermert D., Frise E., Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Seearle S.M.J., Smith E., Shu S., Smutnicki F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* Genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4] SEQUENCE FROM N.A.
 RP Adams M.D., Celtniker S.E., Gibbes R.A., Rubin G.M., Venter C.J.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL

```

DR SMART; SMD0060; FN3; 2.
DR SMART; SMD00408; IgC2; 25.
DR SMART; SMD00406; IgV; 3.
DR SMART; SMD00220; STcC; 1.
DR SMART; SMD00219; TyrKc; 1.
PROSITE; PSS0835; Ig_LIKE; 22.
DR PROSITE; PS00290; Ig_MHC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DON; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding; Immunoglobulin domain; Transferase.
KW SEQUENCE 4463 AA; 497000 MW; DB2E69FCE9B344B CRC64;
SQ

```

-2-

- McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J., Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler E., Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F., Stapleton M., Strong R., Svistak R., Tector C., Tyer D., Wains S.M., Zaveri J.S., Zaveri J.C., Rubin G.M.; RT Sequencing of *Drosophila melanogaster* genome.; Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
- [3]
- SEQUENCE FROM N.A.
Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K., Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D., Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceiriker S.E., Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N., Krommiller B., Marshall B., Millburn G., Richter J., Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield B., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.; RT Annotation of *Drosophila melanogaster* genome.; RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
- [4]
- SEQUENCE FROM N.A.
Adams M.D., Celtniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.; RP Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
- [5]
- SEQUENCE FROM N.A.
RA FlyBase; RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
- [6]
- SEQUENCE FROM N.A.
DR EMBL; AE003008; AAF59087.2; DR GO; GO:0006685; F:calmodulin-dependent protein kinase . . . ; IDA. DR GO; GO:0006687; F:myosin-light-chain kinase activity; IDA. DR InterPro; IPR001599; Ig-like. DR InterPro; IPR007110; Ig-like. DR InterPro; IPR003598; Ig_C2. DR InterPro; IPR003006; Ig_NHC. DR PFam; PF00047; Ig; 19. SMART; SMC00409; Ig; 21. DR PROSITE; PS00408; IGC2_20. DR PROSITE; PS05083; Ig_Like; 19. DR PROSITE; PS00290; Ig_MHC; 1. KW Immunoglobulin domain. SEQUENCE 7210 AA; 8089365 MW; 6B2E7395C07140D2 CRG64;
- Query Match Score 30; DB 5; Length 7210; Best Local Similarity 90.9%; Pred. No. 3.1e+03; Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
- Qy 1 KSAIDLKR 7
Db 6738 KSAIDLKR 6744
- RESULT 8
QBMLD9 PRELIMINARY; ID QBMID9
AC QBMID9; DT 01-OCT-2002 (T-BMBL91; 22, Created)
DT 01-OCT-2003 (T-BMBL92; 22, Last sequence update)
DT 01-OCT-2003 (T-BMBL93; 25, Last annotation update)
DE C318256-PN
GN STRN-MICK OR CG304 OR CG18255
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Dphydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN SP�QUENCE FROM N.A.
RP SP�RAIN-Berkeley; RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celtniker S.E., Holt R.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Ashburner M., Henderson S.N., George R.A., Lewis S.E., Richards S., Rubin G.M., Mungall C.J., Lewis S.E.; OC Drosophila melanogaster (Fruit fly).
OC Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX NCBI_TaxId=7227;
- [1]
- SEQUENCE FROM N.A.
RA Brandon R.C., Rogers Y.-H.C., Blaejaj R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Holt G., Miklos G.I.G., Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Balwair J.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Botsov K.Y., Benos P.V., Berman B.P., Bhakdari D., Brokstein P., Brottier P., Burris K.C., Busam D.A., Butler H., Cadieu B., Centr A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Daventport L.B., Dew I., Dietz S.M., De Pablo B., Delcher A., Deng Z., Dews A.D., Dunn P., Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K., Durbin R.J., Evans-Lista C., Ferraz C., Ferreira S., Fleischmann W., Foster C., Gabreielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.J., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lair Z., Lasko P., Lei Y., Levitsky A.A.C., Li J.J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.A.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Morris J., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon J., Nelson K.A., Nusskern R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Schaefer P., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svartkas R., Tackett C., Turner R., Venter C., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstock G.K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; RT "The genome sequence of *Drosophila melanogaster*."; RL Science 287:2185-2195 (2000). [2]
- RN SEQUENCE FROM N.A.
RA Celtniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y., Banzon J., An H., Baldwin D., Banzon K.Y., Busam D.A., Carlson J.W., Center A., Chapple M., Daventport L.B., Dietz S.M., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D., Ferriera S., Frise E., Gallo R.F., Garg N.S., George R.A., Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J., Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A., McIntosh T.C., Moy M., Murthy B., Nelson C., Nelson K.A., Nunoo J., Paragas V., Park S., Patel S., Pfeiffer B., Photanavong S., Pittman G.S., Puri V., Richards S., Scheeler F., Stapleton M., Strong R., Svirskas R., Tyler C., Tyler D., Venter C., Rubin G.M.; RA Williams S.M., Zaveri J.S., Smith H.O., Venter C.J.C., Rubin G.M.; RA Williams S.M., Zaveri J.S., Smith H.O., Venter C.J.C., Rubin G.M.; RT "Sequencing of *Drosophila melanogaster* genome."; RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases. [3]
- RN SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K., Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D., Hradeccky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D., Tupy J.L., Bergman C., Berman B., Carlson J.W., Celtniker S.E., Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N., Krommiller B., Marshall B., Millburn G., Richter J., Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F., Whittlefield E., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.; RT "Annotation of *Drosophila melanogaster* genome."; RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases. [4]
- RN SEQUENCE FROM N.A.
RA Adams M.D., Celtniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.; RP Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases. [5]
- RN SEQUENCE FROM N.A.
RA FlyBase; RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases. [6]
- RN SEQUENCE FROM N.A.
RA FlyBase; RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases. [7]
- RN SEQUENCE FROM N.A.
RA FlyBase; RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases. [8]
- RN SEQUENCE FROM N.A.
RA FlyBase; RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases. [9]
- RN SEQUENCE FROM N.A.
RA FlyBase; RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases. [10]
- RN SEQUENCE FROM N.A.
RA FlyBase; RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases. [11]
- RN SEQUENCE FROM N.A.
RA FlyBase; RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases. [12]
- RN SEQUENCE FROM N.A.
RA FlyBase; RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases. [13]
- RN SEQUENCE FROM N.A.
RA FlyBase; RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases. [14]
- RN SEQUENCE FROM N.A.
RA FlyBase; RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases. [15]

DR	GO:0004687; F:myosin-light-chain kinase activity; IDA.		
DR	InterPro; IPR003961; FN_III.		
DR	InterPro; IPR008571; FN_-III-like.		
DR	InterPro; IPR007110; FN_-like.		
DR	InterPro; IPR003598; IG_C2.		
DR	InterPro; IPR003066; IG_MHC.		
DR	InterPro; IPR003596; IG_v.		
DR	InterPro; IPR007119; Prot_kinase.		
DR	InterPro; IPR002290; Ser_Thr_Pkinase.		
DR	InterPro; IPR008271; Ser_Thr_Pkinase_AS.		
DR	InterPro; IPR001245; Tyr_Pkinase.		
Pfam	PF00041; fn3; 2.		
DR	Pfam; PF00047; ig; 28.		
DR	Pfam; PF00069; Pkinase; 1.		
DR	ProDom; P000001; Prot_kinase; 1.		
DR	SMART; SMC0408; IgC; 29.		
DR	SMART; SMC0406; IgV; 4.		
DR	SMART; SMC0220; STKC; 1.		
DR	SMART; SMC0219; TYFKC; 1.		
DR	PROSITE; PS00835; IG_LIGS; 27.		
DR	PROSITE; PS00230; IG_MHC; 1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.		
DR	PROSITE; PS00108; PROTEIN_KINASE_ST; 1.		
KW	ATP-binding; Immunoglobulin domain; Transferase.		
SEQ	SEQUENCE 9270 AA; 1036749 MW; QDB82D040B27C2AD CRC64;		
Query Match	Score 30; DB 5; Length 9270;		
Best Local Similarity	90.7%; Pred. No. 4.1e+03;		
Matches	6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;		
Qy	1 KSAIDLKR 7		
Db	6738 RAADLKR 6744		
RESULT 9			
Q832Y5	SEQUENCE FROM N.A. PRELIMINARY; PRT; 139 AA.		
AC	Q832Y5; 01-JUN-2003 (TREMBLrel. 24, Created)		
DT	01-JUN-2003 (TREMBLrel. 24, Last sequence update)		
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)		
DE	Hypothetical protein.		
GN	Enterococcus faecalis (Streptococcus faecalis).		
CC	Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.		
CX	NCBI_TaxID=1351;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=S3 / ATCC 700802;		
RX	MEDLINE=22550857; PubMed=12663927;		
RA	Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.B., Seshadri R., Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F., Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M., Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W., Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M., RT "Role of mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis." Science 299:2071-2074 (2003).		
RL	EMBL; AE016953; AAC081846.1; -.		
DR	TIGR; BF2113; -.		
KW	Hypothetical protein; Complete proteome.		
SEQ	SEQUENCE 139 AA; F757AEC2A159D9C CRC64;		
Query Match	Score 29; DB 16; Length 139;		
Best Local Similarity	87.9%; Pred. No. 92;		
Matches	6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Qy	1 KSAIDLKR 7		
Db	36 KSAIDLKR 42		
RESULT 10			
Q89X71	SEQUENCE FROM N.A. PRELIMINARY; PRT; 186 AA.		
ID	Q89X71; 01-JUN-2003 (TREMBLrel. 24, Created)		
AC	Q89X71; 01-JUN-2003 (TREMBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DB	ATP synthase delta chain.		
GN	ATPH OR BLJ0443.		
OS	Bradyrhizobium japonicum.		
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales.		
OC	Bradyrhizobiaceae; Bradyrhizobium.		
OX	NCBI_TaxID=375;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=S3 / ATCC 700802; PubMed=125997275;		
RX	MEDLINE=22484998; PubMed=125997275;		
RA	Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T., Sasamoto S., Watanabe A., Idezawa K., Iriuchi M., Kawashima K., Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M., Tabata S.;		
RT	"Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110."; RT DNA Res 9:189-197 (2002).		
DR	EMBL; AF005336; BA45708.1; -.		
DR	GO: GO:0005754; P: ATP biosynthesis; IEA.		
DR	InterPro; IPR00711; ArpSynT_OSCP.		
DR	PRINTS; PR00125; ATPASEDELTA.		
DR	TIGR4ANS; TIGR01145; ATP_synt_delta; 1.		
DR	PROSITE; PS00389; ATPASE_DELTA; 1.		
KW	Complete Proteome.		
SEQUENCE	186 AA; 19670 MW; 4125BD0CD9D60750 CRC64;		
Query Match	Score 29; DB 16; Length 186;		
Best Local Similarity	87.9%; Pred. No. 1.2e+02;		
Matches	6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;		
Qy	1 KSAIDLKR 7		
Db	45 ESADLKR 51		
RESULT 11			
Q8BD61	SEQUENCE FROM N.A. PRELIMINARY; PRT; 333 AA.		
ID	Q8BD61; 01-JUN-2002 (TREMBLrel. 21, Created)		
AC	Q8BD61; 01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)		
DR	Thiamine pyrophosphate-dependent dehydrogenases, E1 component alpha subunit.		
DR	ACOA OR TTE0186.		
OS	Thermoanaerobacter tengcongensis.		
OC	Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales.		
OC	Thermoanaerobacteriaceae; Thermoanaerobacter.		
OX	NCBI_TaxID=119072;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MB4 / JCM 11007; PubMed=11997336;		
RX	MEDLINE=21192816; PubMed=11997336;		
RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J., Chen Y., Xue Y., Wu Y., Lai X., Huang L., Dong X., Yang H., Tan H., Chen R., Wang J., Yu J., Ma Y., Ling L.,		
RT	"A complete sequence of T. tengcongensis genome."; RT Genome Res 12:689-700 (2002).		
DR	EMBL; AE01592; AAC03487.1; -.		
DR	GO: GO:016624; F: oxidoreductase activity, acting on the alde... ; IEA.		
DR	GO: GO:000852; F: metallobiotin; IEA.		
DR	InterPro; IPR001017; Dehydrogenase_E1.		



Creation date: 06-10-2004

Indexing Officer: VKHOMRASAVONG - VIVIAN KHOMRASAVONG

Team: OIPEScanning

Dossier: 10240758

Legal Date: 06-09-2004

No.	Doccode	Number of pages
1	CTNF	6
2	892	1
3	1449	1
4	SRFW	1
5	FWCLM	1

Total number of pages: 10

Remarks:

Order of re-scan issued on